Vol. 52 DECEMBER 1961 No. 6

Journal of Educational Psychology

CONTENTS

The Genetic Approach to the Psychology of Thought 275
JEAN PIAGET

Need Achievement Scales as Predictors of Academic Success 282
MERVILLE C. SHAW

The Effects of Direct and Indirect Teacher Influence on Dependent-Prone Students Learning Geometry 286
EDMUND AMIDON AND NED A. FLANDERS

The Influence of Interracial Contact on Social Acceptance in a Newly Integrated School 292
STATEN W. WEBSTER

Teachers' Personality Ratings of High School Pupils 297
H. J. HALLWORTH

Concept Attainment as a Function of Number of Positive Instances Presented 303
EUGENE SHORE AND LEE SECHREST

The Environmental Assessment Technique: A Way to Measure College Environments 308
ALEXANDER W. ASTIN AND JOHN L. HOLLAND

The Desire for Security: An Element in the Vocational Choice of College Men 317
STUART H. BLUM

The Effects of Manner of Speech on Appreciation of Spoken Literature 322
GEORGE M. GLASGOW

Measuring the Value Systems of Education Professors 330
DONALD J. VELDMAN, ROBERT F. PECK, AND CARSON MCGUIRE

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

JOHN B. CARROLL
Harvard University
LEE J. CRONBACH
University of Illinois
FRANCIS J. DIVESTA
Syracuse University
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
DOROTHEA MCCARTHY
Fordham University
CARSON MCGUIRE
University of Texas
PAUL H. MUSSEN
University of California, Berkeley
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
GEORGE G. THOMPSON
Ohio State University

The *Journal of Educational Psychology* publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

ARTHUR C. HOFFMAN, Production Manager
HELEN ORR, Promotion Manager

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-Class postage paid at Baltimore, Md.
Copyright, 1961 by the American Psychological Association, Inc.

Journal of Educational Psychology
1961, Vol. 52, No. 6, 275-281

JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 52 December 1961 Number 6

THE GENETIC APPROACH TO THE PSYCHOLOGY OF THOUGHT[1]

JEAN PIAGET

University of Geneva, Switzerland

From a developmental point of view, the essential in the act of thinking is not contemplation—that is to say, that which the Greeks called "theorema"—but the action of the dynamics.

Taking into consideration all that is known, one can distinguish two principal aspects:

1. The formal viewpoint which deals with the configuration of the state of things to know—for instance, most perceptions, mental images, imageries.

2. The *dynamic* aspect, which deals with transformations—for instance, to disconnect a motor in order to understand its functioning, to disassociate and vary the components of a physical phenomenon, to understand its causalities, to isolate the elements of a geometrical figure in order to investigate its properties, etc.

The study of the development of thought shows that the dynamic aspect is at the same time more difficult to attain and more important, because only transformations make us understand the state of things. For instance: when a child of 4 to 6 years transfers a liquid from a large and low glass into a narrow and higher glass, he believes in general that the quantity of the liquid has increased, because he is limited to comparing the initial state (low level) to the final state (high level) without concerning himself with the transformation. Toward 7 or 8 years of age, on the other hand, a child discovers the preservation of the liquid, because he will think in terms of transformation. He will say that nothing has been taken away and nothing added, and, if the level of the liquid rises, this is due to a loss of width, etc.

The formal aspect of thought makes way, therefore, more and more in the course of the development to its dynamic aspect, until such time when only transformation gives an understanding of things. To think means, above all, to understand; and to understand means to arrive at the transformations, which furnish the reason for the state of things. All development of thought is resumed in the following manner: a construction of operations which stem from actions and a gradual subordination of formal aspects into dynamic aspects.

The operation, properly speaking, which constitutes the terminal point of this evolution is, therefore, to be

[1] This paper, transmitted by Ernest Harms, is the text of an address made by Jean Piaget, from notes, at the New York Academy of Sciences' Conference on the Psychology of Thinking on April 28-29, 1960. The translation was made by Ruth Golbin, who has skillfully succeeded in retaining Piaget's individualistic style of expression.

conceived as an internalized action reversible (example: addition and subtraction, etc.) bound to other operations, which form with it a structured whole and which is characterized by well defined laws of totality (example: the groups, the lattice, etc.). Dynamic totalities are clearly different from the "gestalt" because those are characterized by their non-additive composition, consequently irreversible.

So defined, the dynamics intervene in the construction of all thought processes; in the structure of forms and classifications, of relations and serialization of correspondences, of numbers, of space and time, of the causality, etc. One could think at first glance that space and geometry add to the formal aspect of thought. In this way one conceived of the geometric science in the past, considering it impure mathematics, but applicable to perception and intuition. Modern geometry, since *Le Programme d'Erlangen* by F. Klein, has tended, like all other precise disciplines, to subordinate the formal to the dynamic. The geometries are, indeed, understood today as relying all on groups of transformation, so that one can go from one to the other by characterizing one less general "subgroup" as part of a more inclusive group. Thus geometry too rests on a system of dynamics.

Any action of thought consists of combining thought operations and integrating the objects to be understood into systems of dynamic transformation. The psychological criteria of this is the appearance of the notion of conservation or "invariants of groups." Before speech, at the purely sensory-motor stage of a child from 0 to 18 months, it is possible to observe actions which show evidence of such tendencies. For instance: From 4–5 to 18 months, the baby constructs his first invariant, which is the schema of the permanent object (to recover an object which escaped from the field of perception). He succeeds in this by coordinating the positions and the displacements according to a structure, which can be compared to what the geometricians call "group displacements."

When, with the beginning of the symbolic function (language, symbolic play, imagerie, etc.), the representation through thought becomes possible, it is at first a question of reconstructing in thought what the action is already able to realize. The actions actually do not become transformed immediately into operations, and one has to wait until about 7 to 8 years for the child to reach a functioning level. During this preoperative period the child, therefore, only arrives at incomplete structures characterized by a lack in the notion of combinations and, consequently, by a lack of logic (in transitivity, etc.).

In the realm of causality one can especially observe these diverse forms of precausality, which we have previously described in detail. It is true that a certain number of authors—Anglo-Saxon above all—have severely criticized these conclusions, while others have recognized the same facts as we have (animism, etc.). Yet, in an important recent book (which will appear soon) two Canadian authors, M. Laurendeau and A. Pinard, have taken the whole problem up once again by means of thorough statistics. In the main points they have come to a remarkable verification of our views, explaining, moreover, the methodological reasons for the divergencies among the preceding authors.

At about 7 to 8 years the child arrives at his first complete dynamic structures (classes, relations, and

numbers), which, however, still remain concrete—in other words, only at the time of a handling of objects (material manipulation or, when possible, directly imagined). It is not before the age of 11 to 12 years or more that operations can be applied to pure hypotheses. At this latter level, a logic of propositions helps complete the concrete structures. This enlarges the structures considerably until their disposition.

The fundamental genetic problem of the psychology of thought is hence to explain the formation of these dynamic structures.

Practically, one would have to rely on three principal factors in order to explain the facts of development: maturation, physical experience, and social interaction. But in this particular case none of these three suffice to furnish us with the desired explanations—not even the three together.

Maturation. First of all, none of these dynamic structures are innate, but they form very gradually. (For example: The transitivity of equalities is acquired at approximately 6½ to 7 years, and the ability of linear measure comes about only at 9 years, as does the full understanding of weights, etc.) But progressive construction does not seem to depend on maturation, because the achievements hardly correspond to a particular age. Only the order of succession is constant. However, one witnesses innumerable accelerations or retardations for reasons of education (cultural) or acquired experience. Certainly one cannot deny the inevitable role which maturation plays, but it is determined above all by existing possibility (or limitation). They still remain to be actualized, which brings about other factors. In addition, in the domain of thought, the factors of innateness seem above all limitative. We do not have, for example, an intuition of space in the fourth dimension; nevertheless we can deduce it.

Physical Experience. Experiencing of objects plays, naturally, a very important role in the establishment of dynamic structures, because the operations originate from actions and the actions bear upon the object. This role manifests itself right from the beginning of sensory-motor explorations, preceding language, and it affirms itself continually in the course of manipulations and activities which are appropriate to the antecedent stages. Necessary as the role of experience may be, it does not sufficiently describe the construction of the dynamic structures—and this for the following three reasons.

First, there exist ideas which cannot possibly be derived from the child's experience—for instance, when one changes the shape of a small ball of clay. The child will declare, at 7 to 8 years, that the quantity of the matter is conserved. It does so before discovering the conservation of weight (9 to 10 years) and that of volume (10 to 11 years). What is the quantity of a matter independently of its weight and its volume? This is an abstract notion corresponding to the "substance" of the pre-Socratic physicists. This notion is neither possible to be perceived nor measurable. It is, therefore, the product of a dynamic deduction and not part of an experience. (The problem would not be solved either by presenting the quantity in the form of a bar of chocolate to be eaten.)

Secondly, the various investigations into the learning of logical structure, which we were able to make at our International Center of Genetic Epistemology, lead to a very unanimous

result:[2] one does not "learn" a logical structure as one learns to discover any physical law. For instance, it is easy to bring about the learning of the conservation of weight because of its physical character, but it is difficult to obtain the one of the transitivity of the relationship of the weight:

$$A = C \text{ if } A = B \text{ and } B = C$$

or the one of the relationship of inclusion, etc. The reason for this is that in order to arrive at the learning of a logical structure, one has to build on another more elementary logical (or prelogical) structure. And such structures consequently never stem from experience alone, but suppose always a coordinating activity of the subject.

Thirdly, there exist two types of experiences:

1. The physical experiences show the objects as they are, and the knowledge of them leads to the abstraction directly from the object (example: to discover that a more voluminous matter is more or less heavy than a less voluminous matter).

2. The logicomathematical experience supposes to interrelate by action individual facts into the world of objects, but this refers to the result of these actions rather than to the objects themselves. These interrelations are arrived at by process of abstractions from the actions and their coordinates. For instance, to discover that 10 stones in a line always add up to 10, whether they are counted from left to right or from right to left. Because then the order and the total sum have been presented. The new knowledge consists simply in the discovery that the action of adding a sum is independent of the action of putting them in order. Thus the logicomathematical experience does not stem from the same type of learning as that of the physical experience, but rather from an equilibration of the scheme of actions, as we will see.

Social Interaction. The educative and social transmission (linguistic, etc.) plays, naturally, an evident role in the formation of dynamic structures, but this factor does not suffice either to entirely explain its development, and this for two reasons:

First, a certain number of structures do not lend themselves to teaching and are prior to all teaching. One can cite, as an example, most concepts of conservation, of which, in general, the pedagogs agree that they are not problematic to the child.

The second, more fundamental, reason is that in order to understand the adult and his language, the child needs means of assimilation which are formed through structures preliminary to the social transmission itself —for instance, an ancient experience has shown us that French-speaking children understand very early the expression *"quelques unes de mes fleurs"* [some of my flowers] in contrast to *"toutes mes fleurs"* [all my flowers], and this occurs when they have not yet constructed the relation of inclusion:

Some A are part of all B; therefore

$$A < B$$

In conclusion, it is not exaggerated to maintain that the basic factors invoked before in order to explain mental development do not suffice to explain the formation of the dynamic structures. Though all three of them certainly play a necessary role, they do not constitute in themselves sufficient reason and one has to add to them a fourth factor, which we shall try to describe now.

[2] See *Etudes d'Epistomologie Genetique*, Vol. 7 and 10.

This fourth factor seems to us to consist of a general progression of equilibration. This factor intervenes, as is to be expected, in the interaction of the preceding factors. Indeed, if the development depends, on one hand, on internal factors (maturation), and on the other hand on external factors (physical or social), it is self-evident that these internal and external factors equilibrate each other. The question is then to know if we are dealing here only with momentary compromises (unstable equilibrium) or if, on the contrary, this equilibrium becomes more and more stable. This shows that all exchange (mental as well as biological) between the organisms and the milieu (physical and social) as composed of two poles: (*a*) of the *assimilation* of the given external to the previous internal structures, and (*b*) of the *accommodation* of these structures to the given ones. The equilibrium between the assimilation and the accommodation is proportionately more stable than the assimilative structures which are better differentiated and coordinated.

It is this equilibrium between the assimilation and accommodation that seems to explain to us the functioning of the reversible operations. This occurs, for instance, in the realm of notions of conservation where the invariants of groups do not account for the maturation and the physical experience, nor for the sociolingual transmission. In fact, dynamic reversibility is a compensatory system of which the idea of conservation constitutes precisely the result. The equilibrium (between the assimilation and the accommodation) is to be defined as a compensation of exterior disturbances through activities of the subject orientated in the contrary direction of these disturbances. This leads us directly to the reversibility.

Notice that we do not conceive of the idea of equilibrium in the same manner as the "gestalt theory" does, which makes great use of this idea too, but in the sense of an automatical physical equilibrium. We believe, on the contrary, that the mental equilibrium and even the biological one presumes an activity of the subject, or of the organism. It consists in a sort of matching, orientated towards compensation—with even some overcompensation—resulting from strategies of precaution. One knows, for instance, that the homeostasis does not always lead to an exact balance. But it often leads to overcompensation, in response to exterior disturbances. Such is the case in nearly all occurrences except precisely in the case of occurrences of a superior order, which are the operations of reversible intelligence, the reversible logic of which is characterized by a complete and exact compensation (inverted operation).

The idea of equilibrium is so close to the one of reversibility that G. Brunner, in a friendly criticism of one of our latest books appearing in the *British Journal of Psychology*, proposes to renounce the idea of equilibrium because the notion of the reversibility seems sufficient to him. We hesitate to accept this suggestion for the following three reasons:

First, reversibility is a logical idea, while the equilibrium is a causal idea which permits the explanation of reforms by means of a probabilistic schema. For instance, in order to explain the formation of the idea of conservation, one can distinguish a certain number of successive stages, of which each is characterized by the "strategy" of a progress of compensa-

tion. Now it is possible to show[3] that the first of these strategies (only bearing upon one dimension, to the neglect of others) is the most probable at the point of departure, and further, that the second of these strategies (with the emphasis on a second dimension) *becomes* the most likely—as a function of the result of the first. And, finally, that the third of these strategies (oscillation between the observed modifications upon the different dimensions and the discovery of their solidarity) *becomes* the most likely in the functioning of the results of the preceding, etc. From such a point of view the process of equilibration is, therefore, characterized by a sequential control with increasing probabilities. It furnishes a beginning for causal explanations of the reversibility and does not duplicate the former idea.

Secondly, the tendency of equilibrium is much broader for the operation than the reversibility as such, which leads us to explain the reversibility through the equilibrium and not the reverse. In effect, it is at this level of the obvious regulations and sensory-motor feedbacks that the process of equilibration starts. This in its higher form becomes intelligence. Logical reversibility is therefore conceivable as an end result and not as a beginning and the entire reversibility follows the laws of a semireversibility of various levels.

Thirdly, the tendency to equilibrate does not only explain this final reversibility, but also certain new synthesis between originally distinct operations. One can cite in this regard an example of great importance: the serial of whole numbers. Russell and Whitehead have tried to explain the basic set of numbers through the idea of equivalent classes, without recourse to the serial order. This means that two classes are believed to be equivalent, if one can put their respective elements into a reciprocal arrangement. Only when this relationship relies on the quality of the objects (an A put into relation with an A, a B with a B, etc.) one does not get the quantity. If this relationship is made exclusive of the qualities (an Individual A or B put into relationship with an Individual B or A) then there exists only one way to distinguish the elements from each other. In order not to forget one, or not to count the same twice, one must deal with them in succession and introduce the serial factor as well as the structure of classes. We may then say, psychologically speaking, that the sequence of whole numbers is synthesis between two groupings qualitatively distinct, the fitting of the classes and serialization, and that this synthesis takes place as soon as one excludes the qualities of the elements in question. But how does this synthesis occur? Precisely by a gradual process of equilibration.

On the one hand the child who develops his ideas from numbers is in possession of structures enabling him to fit them into classes (classifications). But if he wants to be exclusive of qualities in order to answer to the question "how many," he becomes unable to distinguish the elements. The disequilibrium which appears, therefore, obliges the child to resort to the idea of order and take recourse to arranging these elements into a lineal row. On the other hand, if the child arranges the elements as 1, 1, 1, etc., how would he know, for instance, how to distinguish the second from the third? This new disequilibrium brings him back to the idea of classification: The "second" is the element which

[3] *Logique et Equilibre*, Vol. 2 of *Etudes d'Epistemologie Genetique*.

has but one predecessor, and the "third" is one that has two of them. In short, every new problem provokes a disequilibrium (recognizable through types of dominant errors) the solution of which consists in a re-equilibration, which brings about a new original synthesis of two systems, up to the point of independence.

During the discussion of my theories, Brunner has said that I have called disequilibrium what others describe as motivation. This is perfectly true, but the advantage of this language is to clarify that a cognitive or dynamic structure is never independent of motivational factors. The motivation in return is always solidary to structural (therefore cognitive) determined level. The language of the equilibrium presents that activity, that permits us to reunite into one and the same totality those two aspects of behavior which always have a functional solidarity because there exists no structure (cognition) without an energizer (motivation) and vice versa.

(Received May 2, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 282-285

NEED ACHIEVEMENT SCALES AS PREDICTORS OF ACADEMIC SUCCESS

MERVILLE C. SHAW
Chico State College

The purpose of the present study was to determine whether or not three of the most widely used scales of need achievement would make appropriate differentiations between groups of high school students matched for ability but differing significantly in mean grade point average. The three scales used in the present research were the Need Achievement scale of the Edwards Personal Preference Schedule, the McClelland Achievement Motivation Test (MAMT), and the French (1956) Achievement scale. Past studies of a similar nature which have made use of need achievement scales have not always been in agreement with each other. McClelland, Atkinson, Clark, & Lowell (1953, p. 238) report a significant positive correlation between their test and grade point average while Lowell (1952) reports nonsignificant correlations between the same two variables.

Another study by Weiss, Wertheimer, and Groesbeck (1960) reported significant correlations between the EPPS Need Achievement scale and the MAMT as well as between each of these scales and the grade point averages of 60 college students in introductory psychology classes. Atkinson (1950) reports a negative but insignificant relationship between MAMT results and grade point average of college students. A study by Morgan (1952) was similar in design to the present study in that two groups of subjects, matched for ability, but differing significantly in grade point averages, were compared on an MAMT. The *t* test yielded a difference between the two groups on the MAMT which was significant at the .02 level.

Procedure

Subjects were selected from the junior and senior classes of a high school of 1,800 students. All students who achieved an IQ above 110 on the California Test of Mental Maturity were considered potential subjects for inclusion in the sample. Students who met the ability criterion and who achieved a grade point average in the previous year of 2.00 (B) or better were classified as achievers. Those who achieved a grade point average of 1.5 or less were classified as underachievers. A total of 18 male achievers and 21 female achievers participated. Twenty male underachievers and 19 female underachievers, likewise, cooperated. All subjects were administered the McClelland Achievement Motivation Test, the Edwards Personal Preference Schedule, and the French Test of Achievement Motivation.

The MAMT was administered to groups through use of an opaque projector under neutral conditions. Cards 7 BM and 8 BM from the TAT and Cards D and H of McClelland's series were the picture stimuli. Subjects responded to the four basic questions used by McClelland. Each MAMT response was scored twice by the author according to the criteria described by McClelland et al. (1953, pp. 107-138). The achievement level of subjects was not known to the scorer at the time protocols were evaluated.

Because of the possibility of low interscorer reliability on the French test, three judges were used in the scoring procedure. A given scoring was considered accurate only when at least two of the three judges responded in identical ways. Judges were not aware of the achievement level of the subjects they were judging.

For purposes of statistical analysis, male and female groups were considered separately. Variance ratio (*F*) and *t* tests were applied to the scores of achiever and underachiever groups on each of the three need achievement measures used in order to establish the existence of homogeneity of

TABLE 1

SIGNIFICANCE OF DIFFERENCE BETWEEN MALE AND FEMALE ACHIEVERS AND UNDERACHIEVERS ON THE CALIFORNIA TEST OF MENTAL MATURITY

Group	CTMM IQ means		F	t
	Achievers	Underachievers		
Male	116.1	115.7	1.108	1.756
Female	117.0	116.7	1.796	1.205

variance and to determine the significance of differences between groups on these measures. In addition, Spearman rank-order correlations between each of the need achievement measures were computed for each of the subgroups in the study in order to determine the extent to which the different achievement measures were measuring the same variable. Finally, an analysis of variance was carried out using the MAMT pictures, achievement level, and sex as differentiating criteria.[1]

RESULTS

The results reported in Table 1 indicate that groups did not differ in performance on the California Test of Mental Maturity. Analysis of the difference between achiever and underachiever groups on each of the three measures of need achievement is reported in Table 2. The only significant difference found was between male achievers and underachievers on the French Need Achievement scale. In none of the other comparisons was the difference between groups found to be significant.

Computation of correlations between each of the three scales for each group used in the study revealed few significant correlations. Table 3 reports these results.

[1] The author would like to express his sincere appreciation to Richard Rankin, Assistant Professor of Psychology, Chico State College, for computing the analysis of variance.

TABLE 2

SIGNIFICANCE OF DIFFERENCES BETWEEN ACHIEVER AND UNDERACHIEVER GROUPS IN THE MAMT, N ACH SCALE OF THE EPPS, AND FRENCH N ACH SCALE

Test	Mean score on criterion		F	t
	Achievers	Underachievers		
MAMT				
Male	5.3	3.5	1.50	1.3503
Female	4.86	3.7	1.46	.8028
EPPS Need Achievement				
Male	16.3	14.6	1.78	1.5889
Female	12.3	12.1	1.17	.1923
French Need Achievement				
Male	8.4	5.9	1.41	2.3640*
Female	7.6	7.2	1.62	.5677

* $p < .05$.

TABLE 3

RANK-ORDER CORRELATIONS BETWEEN EACH OF THE NEED ACHIEVEMENT SCALES FOR EACH GROUP INCLUDED IN THE STUDY

Group	EPPS & MAMT	EPPS & French	MAMT & French
Males			
Achievers	.12	.51*	.25
Underachievers	−.03	.26	.09
Females			
Achievers	.14	−.15	.28
Underachievers	−.23	−.13	.20

* $p < .05$.

The analysis of variance revealed that variation was highly significant (.001) among the four separate pictures of the MAMT but not significant at the .05 level between either achievers and underachievers or between males and females. The same lack of significance was found to exist between all interaction variances as well.

Discussion

Results obtained from the analysis of differences between achievers and underachievers on the MAMT and the EPPS Need Achievement scales would seem to indicate that neither of these instruments would have had predictive value for the groups included in this study. The French Need Achievement scale did reveal a difference significant at the 5% level for male members of the sample.

The single significant correlation was found between the Need Achievement scale of the EPPS and the French Need Achievement scale for male achievers. These three scales not only tend to be poor predictors of academic achievement for this group but are not measuring the same variable.

Results of the analysis of variance of the MAMT indicate that the greatest source of variance exists between pictures rather than between groups of achievers and underachievers, between sexes, or between interaction variances. It would appear that for the groups used in this study the four pictures of the MAMT were not all measuring the same qualities.

It would not be reasonable to conclude on the basis of this single study that present need achievement scales are not predictive of academic attainment, but the fact that other studies also have frequently failed to find any relationship between need achievement scales and academic performance signifies the need for a critical re-examination, not only of need achievement instruments, but of the concept of need achievement itself. The basic assumption underlying the idea of the "need to achieve" is that it is a quantity which some people have in greater degree than do others, and that quantitative measures of this variable will provide an index to the individual's need to achieve in all areas. It would appear that certain fallacies grow out of this asumption. The first fallacy is that the need to achieve is constant for all areas of an individual's life space. The second fallacy is that the need to achieve results in observable production of a socially desirable sort. The third fallacy involved in the basic assumption of the need achievement theory is that felt need to achieve and behavior are congruent. The fact that the intervening variables in the life situation of any given individual may enter into the situation to prevent an individual from meeting any need to achieve he might feel is thus overlooked.

Motives have both force and direction. Present measures of need achievement consider only the former while neglecting the latter. If one is willing to define motivation as the amount of energy expended, then the concept embodied in the idea of "need to achieve" is legitimate. If one believes that motivation cannot be measured in terms of energy output nor amount of production, then the concept of the "need to achieve" will have limited usefulness.

Summary

Subjects matched for ability were classified as achievers or underachievers on the basis of their academic grades. After completing the McClelland Achievement Motivation Test, the French Need Achievement scale, and the Edwards Personal Preference Schedule groups of male and female achievers and underachievers were compared. Results indicated that none of the three need achievement scales differentiated achievers and underachievers significantly, with the exception of the French scale which did differentiate between male achievers and underachievers. The only signifi-

cant correlation was found between the French and the Edward's need achievement scores for male achievers. Analysis of variance indicated that the greatest source of variance existed among the pictures, rather than between groups of achievers and underachievers, or between sexes. Implications of these findings as they related to the concept of need achievement were presented.

REFERENCES

ATKINSON, J. W. Studies in projective measurement of achievement motivation. Unpublished doctoral dissertation, University of Michigan, 1950.

EDWARDS, A. W. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corporation, 1954.

FRENCH, ELIZABETH. Development of a measure of complex motivation. *USAF Personnel Train. Res. Cent. res. Rep.*, 1956, No. TN-56-48.

LOWELL, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1952, **33**, 31–40.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. (Eds.) *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MORGAN, H. H. A psychometric comparison of achieving and nonachieving college students of high ability. *J. consult. Psychol.*, 1952, **16**, 292–298.

WEISS, P., WERTHEIMER, M., & GROSSBECK, B. Achievement motivation, academic aptitude and college grades. *Educ. psychol. Measmt.*, 1960, **20**, 663–665.

(Received February 20, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 286-291

THE EFFECTS OF DIRECT AND INDIRECT TEACHER INFLUENCE ON DEPENDENT-PRONE STUDENTS LEARNING GEOMETRY[1]

EDMUND AMIDON AND NED A. FLANDERS
Temple University *University of Michigan*

Whether or not a particular type of student can learn when he is exposed to a particular style of teaching has interested a number of researchers. Smith (1955) and Wispe (1951) have both shown that when students are classified by the use of personality test data, they respond differently to highly organized versus loosely organized classroom activities in a college remedial reading course (Smith, 1955) and to college lecturing versus group discussion techniques (Wispe, 1951) in freshman sociology. The present project was concerned with dependent-prone eighth grade students who were exposed to consistently direct versus indirect styles of teaching while learning geometry.

Asch (1951), Kagen and Mussen (1956), and Livson and Mussen (1957) have studied the reactions of dependent-prone persons in various kinds of experimental situations. They concluded that dependent-prone individuals are more likely to comply with authority figures and conform to group pressures than the less dependent-prone. Their results suggest that a dependent-prone student might become overly concerned with following the suggestions and directions of a teacher and more dependent on support and encouragement. The present project was designed to find out if these concerns inhibit or enhance the learning of geometry at the age level represented in the eighth grade.

[1] This article is based on the PhD thesis of the first author. It was an adjunctive study of a larger project, the latter supported by a grant from the United States Office of Education.

Procedure

This study employed a laboratory design in order to exercise experimental control of spontaneous behavior. First, the behavior of the teacher was controlled by training a teacher-role player. His statements were classified by an observer to demonstrate that desired differences were great enough for students to notice. Second, those students who scored high on a dependence-proneness test, developed by Flanders, Anderson, and Amidon (1960) were selected for the experimental population. Third, control of learning was accomplished by using pre- and posttests of geometry achievement. And fourth, in half of the experimental groups the basic content material, presented in a tape recording, was so organized that the immediate learning goals were unclear, and in the other half the immediate learning goals were clear.

The four treatments involved were:

Treatment 1. Direct teacher influence: clear goals, 35 dependent-prone students

Treatment 2. Direct teacher influence: unclear goals, 35 dependent-prone students

Treatment 3. Indirect teacher influence: clear goals, 35 dependent-prone students

Treatment 4. Indirect teacher influence: unclear goals, 35 dependent-prone students

The experimental population of 140 dependent-prone students was part of a larger group of 560 students. The larger group was exposed to the preceding four treatments in groups of 20 as part of a larger study. The 560 eighth grade students were selected at random from Minneapolis and St. Paul public schools. The 140 students were the top 25% in each treatment of the larger population according to their scores on the dependence-proneness test.

The students were brought, 20 per session, to a spare room in a public school. First, a pretest of geometry achievement and a test of dependence-proneness were administered. Second, a tape recording was

played introducing the basic concepts of C = D, Distance = Speed × Time, and geometric concepts and formulae concerning inscribed angles. In half of the groups the immediate goals were made clear because the recording explained how this information could be used to solve problems; in the other half, the goals were less clear because the students were warned that they could not be sure how this information could be used. Third, in the direct treatments a teacher gave a 15-minute lecture, with a few questions, explaining the material and illustrating problems that could be solved. In the indirect treatments, a teacher conducted a 15-minute discussion explaining the material and illustrating problems that could be solved. The content coverage was the same in the contrasting treatments. Fourth, the students then had about 15 minutes to practice solving problems at their seats by working on a problem sheet. And fifth, the posttest of achievement was administered. The entire sequence lasted 2 hours.

At appropriate points in this procedure the students' perceptions of goal clarity and their perceptions of the teacher were measured by paper-and-pencil tests. The reliability of both of these scales was estimated by the use of the Hoyt-Stunkard (1952) analysis of variance technique. The estimated reliability of the student perception scale was .64, while the reliability found for the goal perception measure was .93 for the measure administered just before the discussion (Amidon, 1959). Whenever the teacher talked, an observer classified all teacher and student statements according to Flanders' (1960) system of interaction analysis. Later the validity and reliability of the observer's judgments were verified by studying the tape recording that was made of every experimental session.

Results

Control of Teacher Influence and Goal Perception

The manipulation of direct and indirect teacher influence occurred right after the tape recording when the teacher first came in contact with the students. The same role player acted as teacher in all treatments to avoid differences in personality and appearance. The differences between the direct and indirect approach are shown in Table 1 according to the percentage of statements classified into interaction categories.

TABLE 1

Percentage of Tallies in Interaction Categories

Category definition	Treatment			
	1	2	3	4
Teacher talk:				
Praise and encouragement	1.35	1.61	17.04	14.90
Clarification and development of ideas suggested by students	2.48	0.92	15.78	16.10
Asks questions	2.58	1.73	28.07	30.04
Gives own opinions and facts (lectures)	63.10	61.40	13.52	15.97
Gives directions	8.67	10.36	0.28	0.27
Criticizes students	13.03	15.54	1.27	0.94
Student talk	5.07	5.29	16.47	17.17
No one talking	3.49	3.45	7.75	4.69
Total tallies on which the percentage figures are based	889	869	711	746

The figures in Table 1 show that essential differences between the direct and indirect treatments are: the teacher lectures and gives more directions in the direct treatments; he asks more questions and gets more student participation in the indirect treatments; he praises, encourages, and clarifies student ideas more frequently in the indirect treatments; and he criticizes students more frequently in the direct treatments.

The fact that the teacher controlled his behavior successfully and created the two teacher styles is self-evident. A Darwin (1959) chi square analysis of these same interaction data, after they were tabulated in a matrix of sequence pairs, was calculated to test the null hypothesis that there is no difference between interaction data of the direct and indirect combined treatments. The chi square value found was 702.2 (df = 90). This value,

TABLE 2
STUDENT PERCEPTION OF TEACHER INFLUENCE

Group	Mean	F_{obs} between direct and indirect groups
Direct influence	11.74	78.40
Indirect influence	18.84	

transformed to a z score of 24.1 indicates that the differences could have occurred by chance with a frequency of much less than .01.

The reliability of the observers who classified the statements in the live situation was higher than .90. It can be shown that the errors of observation, at this level of reliability, are extremely small compared with the difference shown in Table 1 between the direct and indirect patterns.

After the lectures and discussions and before the period devoted to work sheets, the students responded to a number of opinion items combined into a scale which measured their perceptions of the teacher's behavior. An analysis of variance was made of these scale scores which indicated that the F ratio between the groups subjected to the direct and indirect influence treatments was 78.4 ($df = 1/136$). This was significant at the .01 level of confidence. The mean scores indicated that students in the direct treatments more often saw the teacher as: "telling us what to do," "firm and businesslike," "making plans for us," "critical of our ideas," "talking more than the students," and "not using student ideas or suggestions." In the indirect treatments, students often marked: "finding out what we know," "relaxed and cheerful," "letting us make our own plans," "letting students talk," "using our ideas," and similar perceptions that were the opposite of the direct pattern. The means and F ratio are presented in Table 2.

As expected, an analysis of observation data and of student perceptions of the teacher comparing the clear and unclear goal treatments showed no significant ($p < .05$) difference in teacher behavior. Also, no significant ($p < .05$) interaction effects were found in the analysis of variance.

The interaction analysis data and measures of the students' perceptions of the teacher did show that the differences between the direct and indirect approach did exist as required by the experimental design, were clearly seen by a trained observer, and noticed by the students.

Paper-and-pencil measures of the clarity of goals were made after the playback of the tape recording introducing the basic geometric concepts. Students responded to items such as: "Can you see all the steps necessary to finish your work?" "Right now can you see what you will be doing clearly?" "Can you picture your work so clearly that you could tell when you will be finished?" and similar items that were combined into a scale. An analysis of variance of these scale scores yielded an F ratio of 16.98 ($df = 1/136$, $p < .01$). An inspection of the means indicated that the results were consistent with the intended goal manipulation. The results show that the immediate response of students to the clear and unclear tape recording was significantly different. The means and F ratio are found in Table 3.

Results of the Geometry Achievement

Since achievement in geometry was the fundamental outcome variable analyzed in this study, the postachievement test was subjected to several analyses. The first analysis was the comparison of postachievement

scores between the indirect and direct teacher influence groups. The F ratio found in this analysis was 7.67, which was significant at the .01 level. The means of the indirect teacher influence groups were significantly higher than the means of the direct teacher influence groups on the postachievement measure. In order to reduce unaccounted for error in the analysis of postachievement, two analyses of covariance controlling preachievement scores and intelligence scores were run. The F ratio between direct and indirect teacher influence groups was 10.03 when intelligence was controlled, and 9.62 when pretest scores were controlled. Again, those F ratios were significant at the .01 level, indicating the superiority of the indirect teacher influence group. For each of the analyses of covariance $df = 1/137$.

The analysis of variance and covariance just discussed yielded insignificant ($p > .05$) results when the clear and ambiguous goal perception groups were compared statistically. The interaction of goal perception and teacher influence also did not yield significant ($p < .05$) results. The F ratios found in these analyses are presented in Table 4.

The means of the indirect group were significantly higher than the means of the direct group. This is true when intelligence and preachievement were controlled, and it was also true when they were not controlled. The means of the direct and indirect teacher influence groups are presented in Table 5.

TABLE 3

STUDENT PERCEPTION OF THE GOAL

Group	Mean	F_{obs} between clear and unclear groups
Clear goals	21.40	16.98
Unclear goals	26.13	

TABLE 4

ANALYSIS OF DIFFERENCES IN THE MEANS OF POSTACHIEVEMENT SCORES WITH VARIOUS FACTORS CONTROLLED STATISTICALLY

Source of variation	df	F_{obs}	Null hypothesis
With no measure controlled:			
Interaction	1	—	Not rejected
Teacher influence	1	7.67	Rejected
Goal perception	1	—	Not rejected
Preachievement controlled:			
Interaction	1	2.71	Not rejected
Teacher influence	1	10.03	Rejected
Goal perception	1	—	Not rejected
Intelligence controlled:			
Interaction	1	1.08	Not rejected
Teacher influence	1	9.62	Rejected
Goal perception	1	1.13	Not rejected

TABLE 5

MEANS OF DIRECT AND INDIRECT GROUPS FOR POSTACHIEVEMENT MEASURE

Mean	Not adjusted	Adjusted by pretest scores	Adjusted by intelligence
Direct group	9.24	9.31	9.30
Indirect group	10.82	10.76	10.69

DISCUSSION

The measures of geometry achievement indicate that the dependent-prone students learned more in the classroom in which the teacher gave fewer directions, less criticism, less lecturing, more praise, and asked more questions which increased their verbal participation. This finding takes on added significance when compared

with Flanders' (1960) findings that the total group of 560 students, those scoring high, average, and low on the dependence-proneness tests, failed to show the same significant differences under the same conditions. Moreover, when the independent-prone students (those in the lower 25% on the dependency scale) were compared separately, no differences were found. Compared with students in general, dependent-prone students are apparently more sensitive to the influence pattern of a geometry teacher.

The authors are disposed to interpret these findings in terms of the probable effects of teacher influence on the dependent-prone student. We assume that dependent-prone students are more sensitive to the directive aspects of the teacher's behavior. As the teacher becomes more directive, this type of student finds increased satisfaction in more compliance, often with less understanding of the problem solving steps carried out. Only when he is free to express his doubts, to ask questions and gain reassurance, does his understanding keep pace with his compliance to the authority figure. Lacking this opportunity, compliance alone may become a satisfactory goal and content understanding may be subordinated to the process of adjusting to teacher directives. It is interesting to note that this effect occurs even when the material being learned concerns an orderly, logical system, such as exists in geometry.

One implication of this study is that closer supervision through the use of direct influence, an all too common antidote to lower achievement, may be more harmful than helpful for dependent-prone students.

Summary

The primary purpose of the study described here was to determine the effects of direct vs. indirect teacher behavior and of clear vs. unclear student perception of the learning goal on the achievement of eighth grade geometry students. A specially trained teacher role played both a very direct and a very indirect teacher in a laboratory situation involving 140 eighth grade pupils chosen from a larger population on the basis of high scores on a test of dependency proneness. All students were randomly assigned to one of the following four experimental treatments: direct teacher influence with clear goals, direct teacher influence with unclear goals, indirect teacher influence with clear goals, and indirect teacher influence with unclear goals.

Students in the various classifications were then compared on the basis of pre- and postachievement tests in geometry. No differences were found between the clear goal and unclear goal treatments, indicating that in this study, at least, achievement of dependent-prone students was not affected by perception of the learning goal. An analysis of the direct and indirect treatments indicated that the children taught by the indirect teacher learned more than did the children taught by the direct teacher.

The results of this study take on additional meaning when compared with the results of Flanders (1960) using the same experimental design. Flanders found no differences (among the four experimental conditions) on the total group of 560 students who ranged from very high to very low on the dependence scale. Apparently dependent-prone students are more sensitive to types of teacher influence than are independent-prone students or students who make average scores on the test for dependence proneness.

REFERENCES

AMIDON, E. J. Dependent-prone students in experimental learning situations. Unpublished doctoral dissertation, University of Minnesota, 1959.

ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men.* Pittsburgh; Carnegie Press, 1951.

DARWIN, J. H. Note on the comparison of several realizations of a Markoff chain. *Biometrika*, 1959, **46**, 412–419.

FLANDERS, N. A. Teacher influence on pupil attitudes and achievement. Final Report; 1960; University of Minnesota; Project 397; United States Department of Health, Education, and Welfare; Cooperative Research Program; Office of Education.

FLANDERS, N. A., ANDERSON, J. P., & AMIDON, E. J. Measuring dependence proneness in the classroom. *U. Minn., Bur. Educ. Res. res. Memo.*, 1960, No. Ber 60–6.

HOYT, C., & STUNKARD, C. L. Estimation of test reliability for unrestricted item scoring methods. *Educ. psychol. Measmt.*, 1952, **12**, 756–758.

KAGEN, J., & MUSSEN, P. Dependency themes on the TAT and group conformity. *J. consult. Psychol.*, 1956, **20**, 19–27.

LIVSON, N., & MUSSEN, P. The relation of control to overt aggression and dependency. *J. abnorm. soc. Psychol.*, 1957, **55**, 66–71.

SMITH, D. E. P. Fit teaching methods to personality structure. *High sch. J.*, 1955, **39**, 167–171.

WISPE, L. C. Evaluating section teaching methods in the introductory course. *J. educ. Res.*, 1951, **45**, 161–186.

(Received March 13, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 292-296

THE INFLUENCE OF INTERRACIAL CONTACT ON SOCIAL ACCEPTANCE IN A NEWLY INTEGRATED SCHOOL[1]

STATEN W. WEBSTER[2]

University of California, Berkeley

Does interracial contact lead to improved attitudes and increased social acceptance among the participants? This question has been the subject of a number of investigations, most of which have involved white and Negro adult subjects.

Data reported in these studies have been contradictory. A number of investigations (Brophy, 1945; Deutsch & Collins, 1952; Neprash, 1952; Wilner, Walkley, & Cook, 1952) have indicated the successful influence of interracial contact in reducing prejudice. However, a number of negative findings have also been reported (Mussen, 1950; Rose, 1947; Young, 1932).

Those who argue against the positive effects of contact cite the restrictive influence of racial and ethnic group norms upon individuals in such situations. The existence of such frames of reference has been clearly documented by Bogardus (1959). While the impact of racial or ethnic group membership upon the individual has been established, there is also evidence that under certain conditions individuals and groups do modify their interracial attitudes, feelings, and behaviors. An analysis of successful contact situations seems to reveal three important requisite factors: equal status of the participants in the situation, contact of sufficient duration and intimacy, and the sanction or support of a higher authority.

[1] This study is abstracted from S. W. Webster's doctoral dissertation written in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of California, Berkeley, 1960.

[2] Now Associate Research Director of the project, Studies in the Dynamics of Administrative Interaction, at the School of Education, University of California, Berkeley.

PROCEDURE

Problem

In view of the dearth of contact research involving public school settings, this investigation sought to study the effects of interracial contact in a newly integrated northern California junior high school.

Specifically, two major questions were investigated: Does interracial contact, over a 6-month period, between white and Negro subjects, lead to increased expressions of social acceptance of one group for the other? Are precontact social acceptance scores (as measured by a social acceptance scale) associated with the eventual making of interracial friendships?

Setting. The school board of a high school district in the San Francisco Bay Area reached a decision in the spring of 1959 to effect certain changes in the boundaries surrounding its junior and senior high schools. These realignments were made necessary by overcrowded conditions in some schools, and the unequal distribution of children in terms of socioeconomic status.

Prior to the boundary modifications, which became effective in September 1959, Jackson Junior High School[3] was populated by an all white student body. The composition of the school population was altered greatly by the boundary changes, with some 300 Negro children being integrated into the school's present total of some 1,400 students. This school, its feeder elementary schools, and its surrounding communities served as the setting for this study. While there were some socioeconomic differences between the white and Negro communities, a study of the personal histories of the children revealed these to be minor in nature.

Subjects. When the new school boundaries were announced late in the spring, it became

[3] Actual school names are not used in this report.

obvious that most children in two elementary schools would be attending, in the fall, a truly integrated school for the first time. Walker Elementary School, which was all white, sent 60 sixth grade graduates to Jackson Junior High in September. From the almost all Negro Newell School came 44 transfers. (This school contained, in addition to some 400 Negro students, 17 children of Indian and Mexican-American extraction.)

A socioeconomically similar control group comprised of 55 whites and 53 Negroes was also selected. An all white community and junior high school, in a neighboring city, provided the white control subjects. The Negro control subjects were selected from an almost all Negro junior high school in another nearby city; with the exception of 9 Mexican-American and white children, the latter school's student population of 450 was composed entirely of Negroes.

Hypotheses. Four hypotheses were investigated in this study; they were as follows:

Hypothesis I: White and Negro subjects who have experienced intergroup contact in a school situation will express significantly more social acceptance of each other after a 6-month period of time.

Hypothesis II: The postcontact social acceptance scores of the white and Negro experimental subjects will display significantly more out-group acceptance than will the noncontact scores of the control subjects.

Hypothesis III: The precontact social acceptance scores of the experimental subjects will be significantly associated with their later friendship choices.

Hypothesis IV: The retest social acceptance scores of the experimental subjects will be significantly associated with their interracial friendship choices.

These hypotheses are interrelated. Hypotheses I and II seek to measure the impact of contact upon the experimental subjects, in light of the data on the control subjects. Hypotheses III and IV are designed to seek additional data on the effects of contact and on the validity of the social acceptance scale used to gather these data.

Method

The test-retest design was deemed most appropriate for this type of study. Both experimental groups were tested initially in June 1959, while they were still attending their particular elementary schools, and were retested at Jackson School in March 1960. Because of the time factor, it was impossible to test the control groups in spring 1959. The failure to obtain these data is, admittedly, a weakness of the study. However, since their elementary and junior high schools were relatively homogeneous in racial composition, it seems valid to assume that their single test scores (spring 1960) on a social acceptance scale (see below) would not differ significantly from the precontact scores of their matched experimental groups.

All testing sessions were conducted by members of the subject's racial group. To avoid the possibility of inhibiting the responses of the children, they were told that the questionnaires were anonymous. However, although the experimenter entertained some reservations about the procedure, invisible ink markings were used on the social acceptance scale to identify the respondents.

To measure social acceptance, a nine-item social acceptance scale was constructed. The respondent was asked to rank his willingness to accept members of four groups (Chinese, Japanese, Mexicans, and whites or Negroes) in nine social situations. The social situations ranged from the less intimate (e.g., working in the same factory) to the very intimate (e.g., living next 'door). Scale scores from 0 to 9 were possible, with the higher scores representing the greater acceptance. The scalability of the instrument was established in keeping with the criteria suggested by Guttman (1950). The coefficient of reproducibility was .92.

In addition to the social acceptance scale, a brief sociometric friendship questionnaire was used. The subjects were asked to indicate their six best friends in the school and to give the ethnic or racial backgrounds of these friends. These data were obtained after the retest of the experimental subjects.

To test the above hypotheses, it was necessary to validate the assumption that both experimental and control groups were initially similar in levels of social acceptance. The two-tailed application of the Kolmogorov-Smirnov two-sample test (Siegel, 1956) was used in this analysis. Since the KS test is sensitive to differences in central tendency, dispersion, and skewness, it seemed especially appropriate for this purpose. The comparison of the experimental subjects' precontact scores and the single test scores of the control subjects revealed no significant differences.

Results

Table 1 contains a summary of the basic test data on the experimental

TABLE 1
TEST AND RETEST SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON A SOCIAL ACCEPTANCE SCALE

Subjects	Group	N	Range	$\bar{X}$	Mode	SD
Experimental:						
Pretest (June 1959)	White	60	0–9	1.85	0	2.32
	Negro	44	0–9	6.05	9	2.86
Posttest (March 1960)	White	60	7–0	1.48	0	2.02
	Negro	44	9–0	6.32	9	2.61
Control: (March 1960)	White	55	0–9	2.54	0	2.24
	Negro	53	0–9	5.98	6	3.97

and control groups. The reader will note the great discrepancies between the scores of the Negroes and the whites. Bogardus (1959) has reported similar distributions in one of his studies of social distance.

Hypothesis I. The Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) was used to test the first hypothesis.[4] This nonparametric test makes use of comparisons between pairs of scores and takes into consideration both the magnitude as well as the direction of the differences between the sets of scores. Change for the white subjects was in the negative direction and significant ($z = 3.96$, $p < .05$). Change for the Negro subjects was positive and significant ($z = 2.12$, $p < .05$). Thus, the predicted hypothesis was confirmed only in the case of the Negro subjects.

An equal number of Negro subjects changed in each direction. However, the magnitude of change was greater for those subjects whose gains were positive.

Hypothesis II. The testing of Hypothesis II involved comparing the postcontact scores of the experimental groups with those of their controls, in an attempt to see if the contact situation had produced significant differences among the groups.

The one-tailed application of the KS test revealed a significant difference between the scores of the white experimental and control groups ($\chi^2 = 7.23$, $df = 2$, $p < .05$); however, this difference was in the nonpredicted direction. The obtained difference between the scores of the Negro experimental and control subjects was not significant ($\chi^2 = 2.16$, $df = 2$, $p > .05$).

Hypothesis III. This hypothesis is based on a suggestion by Neprash (1952). He speculated that the making of out-group friendships in a contact situation may be simply a matter of favorable precontact attitudes toward that group and not a result of contact. To test this assumption the precontact scores of the experimental subjects were dichotomized at the median and a comparison was made of the number of out-group friendships indicated by each group. The χ^2 test was used in the analysis and the findings were not significant. The median, in this case, happened to fall close to the derived zero intensity point (discussed later).

These findings seem to indicate that, in this study, precontact attitudes played little or no part in the willingness to make out-group friendships in the contact situation. The making of interracial friendships, it appears, was related to the effects of contact itself.

Hypothesis IV. It was predicted in Hypothesis IV that the subjects' retest social acceptance scores would be significantly associated with their reported interracial friendships. In testing this assumption, use was made of the intensity measure suggested by Guttman (1954) and his associates.

While the method of deriving this score will not be fully discussed, a

[4] The .05 level of significance was used in testing all hypotheses.

brief description will be given. The zero or low intensity point is a measure that seeks to establish a psychologically meaningful point of low intensity on an attitude scale. Through the use of the extreme responses of the subjects to each scale item (Answer Categories 1 or 4), it is possible to trichotomize the respondents into three general groups: intensely positive subjects, intensely negative subjects, and indifferent or low intensity subjects.

In the case of the social acceptance scale used in this study, the zero intensity point was found to be located at the scale score of four.

The subjects were dichotomized on the basis of their intensity scores, and an analysis was then made of the number of out-group friendship choices reported by each group. The obtained result was significant ($\chi^2 = 10.56$, $df = 1$, $p < .01$) and Hypothesis IV, was thus confirmed.

Discussion

The positive effects presumably associated with interracial contact were not confirmed in this study. What, then, are some factors which seem to account for the obtained results?

Although the three-part criterion presented earlier in this paper was met in this contact situation, it is possible that only more general and intimate contact between interracial groups (Brophy, 1945; Deutsch & Collins, 1952; Wilner et al., 1952) results in significant modifications of attitudes. It should be remembered that the subjects, while attending the integrated school, continued to live in racially closed communities. It is possible that integration within the total social environment is necessary, for the attitudes and values which can be developed in the school setting are probably contradicted by the realities of the restricted community.

The length of the contact situation in this instance could be another factor that partially explains the failure of the contact hypotheses. A 6-month period of time could very well be inadequate to allow for the necesary intergroup adjustment of the subjects. Although this period of contact was longer than that reported in a number of studies, subsequent investigation may show that the initial assumptions associated with contact experiences are valid.

The social acceptance scores of both white and Negro subjects provide another possible explanation. Social distance (acceptance) research by Bogardus (1959) and others has documented the existence of a hierarchy of racial preferences. These normative frames of reference appear to be held both commonly and consistently by members of various racial or ethnic groups. Traditionally, whites rate Negroes in their preferences low on a social distance scale, while Negroes tend to rate whites quite highly. The data obtained in this study reflect these phenomena. While social distance measures have proven to be both reliable and valid, it is possible that they are not sensitive enough to measure less noticeable changes within individual attitude structures. Methods and instruments which assess interracial attitudes with greater depth and precision would seem to be needed.

An interesting finding of the study was the significant association of interracial friendship choices with postcontact social acceptance scores but not with pretest scores. This finding seems to suggest caution in accepting the responses of subjects to social distance and other types of attitude measures as predictors of behavior. Many subjects whose precontact responses were favorable or highly accepting of the other group failed to

make eventual interracial friendships, and in a number of cases completely reversed their scale scores from high acceptance (scores of 7–9) to high rejection (0–2).

The data obtained in this study seem to suggest a critical reappraisal of the contact hypothesis. Factors external to the setting of the interaction as well as those associated with racial and ethnic group membership must be considered.

Summary

This study tested three major hypotheses relating to interracial contact, in a California junior high school, between white and Negro subjects. It was held that (*a*) a school contact situation of 6-months' duration would lead to increased interracial social acceptance by members of the two racial groups, (*b*) that a comparison between experimental subjects who experienced interracial contact and similar control subjects who did not would reveal greater interracial acceptance among the former group, and (*c*) that precontact social acceptance scores of the experimental subjects would be associated with out-group friendship choices made in the contact setting.

The test-retest design was used with 60 white and 44 Negro subjects, and 55 white and 53 Negro control subjects. A single administration of the social acceptance scale took place with the control subjects. All tests were administered by adults of the particular subjects' racial groups.

Contact had a negative effect upon the white subjects; they became significantly less accepting of Negroes. The findings were inconclusive in the case of the Negro subjects, but did tend to indicate that change was greater in the direction of more acceptance of whites. Precontact social acceptance scores were not found to be associated with later friendship choices. However, postcontact acceptance scores were associated with actual out-group friendship choices at the .01 level of confidence.

REFERENCES

Bogardus, E. Race reactions by sexes. *J. Sociol. soc. Res.*, 1959, **43**, 439–441.

Brophy, I. The luxury of anti-Negro prejudice, *Publ. opin. Quart.*, 1945, **9**, 456–466.

Deutsch, M., & Collins, Mary E. The effects of public housing upon interracial attitudes. In G. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (Rev. ed.) New York: Holt, 1952. Pp. 582–593.

Guttman, L. The basis of scalogram analysis. In S. Stauffer, E. A. Suchman, P. F. Lazarsfeld, Shirley A. Star, & J. A. Clausen (Eds.), *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. Pp. 60–90.

Guttman, L. Principal components of scalable attitudes. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. Glencoe, Ill.: Free Press, 1954. Pp. 229–233.

Mussen, P. H. Some personality and social factors related to changes in children's attitudes toward Negroes. *J. abnorm. soc. Psychol.*, 1950, **45**, 423–441.

Nepbash, J. Racial group contacts and social distance. *Phylon*, 1952, **16**, 207–212.

Rose, A. M. *Studies in reduction of prejudice*. Chicago: American Council on Race Relations, 1947.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Wilner, D. M., Walkley, R. R., & Cook, S. W. Residential proximity and intergroup relations in public housing projects. *J. soc. Issues*, 1952, **8**, 45–69.

Young, D. *American minority peoples*. New York: Harper, 1932.

(Received April 3, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 297-302

TEACHERS' PERSONALITY RATINGS OF HIGH SCHOOL PUPILS

H. J. HALLWORTH

Department of Education, University of Birmingham, England

Recent contributors to this journal have indicated the need for research into teachers' ratings of their pupils, a need which is equally pressing in Europe as in America. The present article describes part of such a research which was carried out from the Department of Education in the University of Birmingham, England. It therefore concerns a sample drawn from a different educational tradition from those studied, for example, by Ryan (1958) and Holland (1959). The results, however, are remarkably comparable to theirs and, with one exception, discrepancies appear to be due entirely to differences in measuring instruments and statistical procedures.

It is hoped that comparative data such as this will be of interest to American workers both in affording confirmation of their findings and in suggesting small differences which may prove illuminating.

Method

Subjects

At the age of 11+ years most English children pass from primary to secondary schools. Approximately 20% with the highest intelligence and attainment are allocated to grammar schools; most of the remainder pass into secondary modern schools.

Ratings were obtained for boys in first year (11+ years) and fourth year (14+ years) classes of six single-sex secondary schools in an industrial region near Birmingham. The schools were selected in three matched pairs, each pair comprising a boy's grammar school and a boy's secondary modern school within one locality. The subjects were therefore divided according to age and school type into four groups, each distributed over three schools and containing approximately 112 boys.

It is commonly accepted that there are sex differences in certain personality traits: boys are assumed to be more self-assertive, and girls more cooperative with adults. Such differences have frequently been demonstrated in mean scores obtained on personality inventories (e.g., Castaneda, McCandless, & Palermo, 1956). Groups comprising members of both sexes may consequently be expected to yield spuriously high correlations between certain personality traits. It was therefore considered necessary to obtain ratings for groups consisting of pupils of like sex only. However, in the complete research of which part is reported in this article, ratings were obtained for parallel groups of girls and yielded results comparable to those given below.

Similarly, it may be claimed that some personality traits are more typical of one age than of another. Ratings obtained for a group with a wide age range may therefore have spuriously high intercorrelations. It was therefore considered necessary to obtain ratings for several classes of boys all of the same age level, and to combine these into one group. This was done for two age levels and for two types of school in order to determine the effect of age and school type upon the intercorrelations obtained.

Measures

Each class teacher rated his pupils on 12 personality traits, using a graphic five-point scale. In order to avoid spuriously high intercorrelations from another source, he was requested to complete all ratings on one trait at a time. For each trait there was provided a brief description for each of the extreme ends of the scale, and a two-word description of each of the five points on the scale. The traits were as follows:

1. Emotional stability
2. Trustworthiness
3. Persistence
4. Cooperation with teachers
5. Cheerfulness
6. Sense of humor
7. Sociability
8. Self-assertion–submission
9. Maturity
10. Popularity
11. Confidence-anxiety
12. Spontaneity-withdrawal

When several teachers rate different children and the ratings on different traits are then intercorrelated, there is yet another possible source of spurious correlation: some teachers give high ratings on all traits, some give low ratings. In the present research therefore, no attempt was made to use ratings made by different teachers when obtaining correlations between traits. It was accepted that neither mean nor standard deviation could be assumed to be comparable for ratings made by different teachers or, indeed, for ratings made by one teacher on different traits. Teachers were therefore requested, when making ratings on any one trait, to take the mean of their own class as the mean of their rating scale. They were also requested to distribute the ratings according to the percentages of the normal curve, a graphic model being provided for this purpose.

Analysis

The data was analysed with the aid of an electronic computer. A correlation matrix was first obtained for the ratings made on the 12 traits by each teacher for the boys in his own class. Within each of the four groups the correlation matrices were then combined, using z scores, to obtain a matrix of average intercorrelations between traits for each group. A principal components analysis was then performed on each of the four combined matrices.[1]

Results

Across each of the four analyses the first two components were essentially similar and accounted for approximately 60% of the total variance. Further components invariably accounted for less than 10% of the total variance and, since they had little similarity across the analyses, will not be considered in this article. There is general agreement regarding the large error variance in teachers' ratings. However, when comparable factors are obtained from four different analyses, this suggests a certain regularity in the manner in which ratings are made and justifies further examination.

The first two components for each of the four analyses are shown in the accompanying diagrams, Component I being shown along the x axis and Component II along the y axis. The percentage of the total variance extracted by each component is indicated.

In the diagram for first year grammar school boys, a group of ratings falls close to each of the main axes. Those falling near the axis for Component I are cheerfulness, sense of humor, sociability, self-assertion, and spontaneity, each with approximately 64% of its variance accounted for by this one dimension. The ratings falling approximately along the axis for Component II are emotional stability, trustworthiness, persistence, cooperation with teachers, and maturity, each with between 38% and 66% of its variance accounted for by this dimension. We may name the first dimension "extraversion" and the second "reliability and conscientiousness." Two other ratings, namely popularity and confidence, lie close together and between these dimensions.

The diagram for fourth year grammar school boys has the same two clusters of ratings, albeit they are in this case less compact. A slight orthogonal rotation would bring them onto the main axes. Again the ratings on confidence and popularity lie together and between the two clusters.

The two diagrams for secondary modern school boys appear at first sight to show the well-known "halo-effect" which occurs when ratings on a number of traits are made of an individual. All ratings have a heavy loading on Component I. For first year boys, however, the same two clusters may be seen, although the extraversion traits are more widely spread. In the analysis for fourth year boys, one

[1] Acknowledgements are due to W. L. B. Nixon, of the University of London Computer Unit, whose Mercury computer principal components (analysis) program was used. The program for the combination of correlation matrices, using z scores, was developed by the author in conjunction with the Computer Unit.

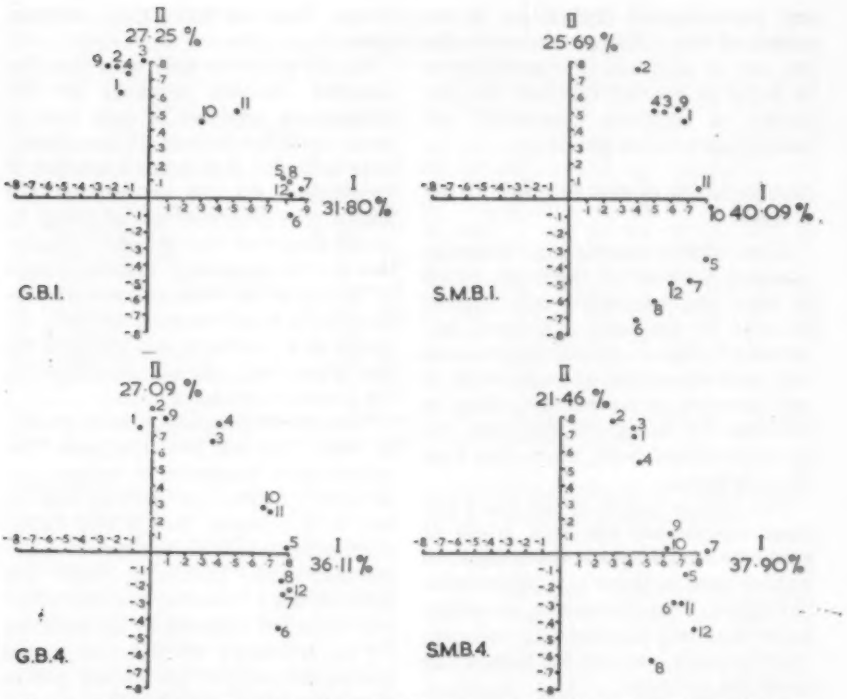

FIG. 1. Principal components analyses for four groups: Loadings of twelve ratings on two components. (I = first component, II = second component. The precentage of variance extracted by each component is shown. GB_1 = grammar school boys Year 1, GB_4 = grammar school boys Year 4, SMB_1 = secondary modern school boys Year 1, SMB_4 = secondary modern school boys Year 4. 1 = emotional stability, 2 = trustworthiness, 3 = persistence, 4 = cooperation with teachers, 5 = cheerfulness, 6 = sense of humor, 7 = sociability, 8 = self-assertion-submission, 9 = maturity, 10 = popularity, 11 = confidence-anxiety, 12 = spontaneity-withdrawal.)

cluster is made by the traits of trustworthiness, persistence, emotional stability, and cooperation. The extraversion traits are again scattered and now include among them the ratings of confidence, popularity, and maturity.

DISCUSSION

Grammar and Secondary Modern School Groups

It is possible that the greater degree of apparent halo effect for the two secondary modern school groups is due to psychologically irrelevant factors in the data or in the method of analysis. Further research is necessary to determine the matter, and is currently in progress. A consistent difference related to school type appears, however, to demand a psychological explanation.

There is no reason to suppose that the secondary modern school teachers had less acquaintance with their pupils and therefore less opportunity to differentiate their ratings. Nor is there any reason to assume that their ratings were in any other way made under less favorable conditions.

It would appear, therefore, that if

any psychological explanation is required of the difference between the two sets of analyses, it is probably to be found in the way in which the two groups of teachers "perceived" or thought about their pupils.

Interpretation of the Principal Components

Ryan (1958) reports high intercorrelations between all the traits which he used and concludes that "teachers rate on the basis of general impression." This is, admittedly, a common and reasonable conclusion. It is not, however, an invariable finding, as is shown by the above diagrams for grammar school boys, where litle halo effect is found.

The interpretation of the two principal components for each group of grammar school boys is quite clear. In neither case is there any appreciable halo effect. Two dimensions, one of extraversion, one of reliability and conscientiousness, account for three-fifths of all the variance.

The problem arises when we consider the secondary modern school groups.

If we interpret the two principal components as they stand, we shall label Component I "halo effect." Component II is, however, difficult to name, and here lies the difficulty: It must be described in terms of a scale running from cooperation, trustworthiness, persistence, and emotional stability at one end, to sociability, cheerfulness, spontaneity, self-assertion, and sense of humor at the other. Such a scale does not make good psychological sense. Moreover, the suggestion that the clusters which we have named "reliability and conscientiousness" and "extraversion" are independent dimensions for grammar school groups, but opposite ends of a single scale for secondary modern school groups, does not make good common sense.

At this point we may note that the factorial solution provided by the components analysis is only one of many possible solutions. It has already been indicated that a small rotation is appropriate for the analysis of the fourth year grammar school group. It is now suggested that the obvious solution for the secondary modern groups is to rotate the two components orthogonally until the apparent halo effect is at a minimum and we have the two dimensions already obtained for the grammar school groups.

That this is practicable may readily be seen from the two diagrams. The extraversion dimension is indeeed not as clearly defined as could be desired, and it is possible that a still better description would be in terms of a circumplex model (Schaefer, 1959). The rotation does, however, produce better psychological meaning in the analyses for the secondary modern groups, and makes the analyses for all four groups comparable.

Traits and Dimensions

If we accept this solution, there remains to consider the difference between the present research and certain American work, such as that already quoted, which has been interpreted in terms of halo effect only. The difference is almost certainly due in part to different procedures employed in obtaining ratings. Table 1 gives the correlation matrices for the two first year groups, and it is readily apparent that the correlations are lower than those obtained by either Ryan or Holland. The most significant difference lies, however, in the nature of the traits used. Ryan, for example, used eight traits: seriousness of purpose, industry, initiative, influence, concern for others, responsibility, self-control,

TABLE 1

THE CORRELATION MATRICES FOR TWELVE TEACHER RATINGS FOR GRAMMAR SCHOOL BOYS YEAR 1 AND SECONDARY MODERN SCHOOL BOYS YEAR 1

Trait	1	2	3	4	5	6	7	8	9	10	11	12
1. Emotional stability		46	40	27	−03	−15	−08	−08	40	09	37	−23
2. Trustworthiness	55		56	53	−03	−16	−16	−13	49	28	21	−15
3. Persistence	54	55		57	12	−11	−06	09	58	21	32	02
4. Cooperation with teachers	51	59	37		04	−16	−01	−12	55	24	14	−02
5. Cheerfulness	43	11	25	20		71	58	45	−18	21	37	56
6. Sense of humor	−07	−35	−11	−13	67		57	58	−31	21	30	55
7. Sociability	19	00	10	23	75	56		66	−09	24	44	74
8. Self-assertion–submission	08	−28	02	−08	56	42	66		00	30	46	59
9. Maturity	61	62	58	49	30	−15	15	11		26	27	−06
10. Popularity	50	31	37	44	66	36	59	32	44		27	15
11. Confidence-anxiety	51	25	52	24	48	15	42	38	56	53		31
12. Spontaneity-withdrawal	21	−08	07	−05	56	49	57	59	18	42	46	

Note: Correlations for grammar school boys Year 1 are above the diagonal. Correlations for secondary modern school boys Year 1 are below the diagonal.

neatness and personal appearance. All would appear to fall on, or close to, the dimension of reliability and conscientiousness. The same comment applies, in general, to the 12 ratings used by Holland except, possibly, popularity, social leadership, and physical vigor. In neither case was any typical set of extraversion traits used. It would therefore be impossible to obtain two dimensions similar to those obtained in the present study, and the most comparable result would be the halo effect that was actually obtained in each case.

In the present research, ratings on the traits of popularity and confidence tend to fall together and between the two main dimensions. Here, possibly, is the beginning of a third dimension which is not, however, orthogonal to the other two. There is no reason to suppose that yet other dimensions could not be identified. It is, however, apparent that they would be of less importance both practically and theoretically. It is equally apparent that the dimension of reliability and conscientiousness, already so well identified in earlier articles in this journal, is of greater importance to teachers than is the dimension of extraversion.

Ratings of Maturity

The rating of "maturity" is of particular interest in the present research, since in three analyses it is part of the dimension of reliability and conscientiousness, and in the analysis for fourth year secondary modern school boys it is closer to the extraversion traits. An interpretation of this difference appears to demand some acquaintance with secondary education in England.

Grammar schools in England are essentially academic schools and grammar school teachers value most highly in their pupils those traits which are associated with academic success. Both these teachers and, apparently, teachers of first year secondary modern school boys, regard as mature those pupils who are most trustworthy, persistent, emotionally stable and cooperative: in other words, the children who are most mature in their school work.

Teachers of fourth year secondary modern school boys associate maturity in their pupils more closely with those

traits which may be immediately useful in life outside school. Very few of these boys could be expected to remain at school beyond the end of their fourth year. It appears that their teachers were already assessing their maturity in terms of the social qualities they would require upon leaving school.

The implication is that ratings of maturity have a large "evaluative" element. It is probable that most teachers, as Holland found, measure maturity in relation to the dimension of reliability and conscientiousness. There appear to be, however, circumstances in which teachers relate maturity to other traits which they consider more relevant for their pupils.

Conclusion

The results of this research indicate that the extent to which trait names are interchangeable depends upon the traits with which we are dealing. If we refer to those traits with which teachers are commonly concerned and which in this article have been designated as reliability and conscientiousness, then indeed they are interchangeable as Ryan suggests; to measure one of them, such as persistence or emotional stability, would be effectively to measure all. If, however, we also refer to such traits as sociability, then there is a case for maintaining that here is a quite different dimension. If still other traits were used, other dimensions again might appear. "Popularity" and "confidence" tend to be synonymous, and relatively independent of either of the two dimensions obtained in this research. "Maturity" changes its meaning, as if teachers always ask themselves: "Maturity for what?"

In so far as we obtain a halo effect we may say that when a teacher is rating one of his pupils on a number of traits he tends to ask himself one question only: "How does he get on with me?" or "How much do I approve of him?" In effect, it is suggested in this article that, if we look more closely at ratings on a variety of traits, we may reasonably conclude that teachers ask themselves not one question, but two, namely, "How does he get on with me?" (reliability and conscientiousness); and "How does he get on with other boys?" (extraversion).

It is apparent that the answers to these two questions alone will go far to explain teachers' ratings on all the traits used in this research.

The question whether they may be described as referring to two teachers' stereotypes will be dealt with in a later article.

Summary

Teachers in six boys' secondary schools in England rated their pupils on 12 personality traits, and a correlation matrix was obtained for each teacher. Matrices were combined for classes of different age and school type into four groups. In contrast to the usual finding of a "halo" effect, two dimensions were obtained in each case and were identified as "reliability and conscientiousness" and "extraversion." It is suggested that when rating pupils on a variety of traits, teachers may be described as asking themselves two questions: "How does he get on with me?" and "How does he get on with other boys?"

REFERENCES

Castaneda, A., McCandless, B. R., & Palermo, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, **27**, 317–326.

Holland, J. L. Some limitations of teacher ratings as predictors of creativity. *J. educ. Psychol.*, 1959, **50**, 219–223.

Ryan, F. J. Trait ratings of high school students by teachers. *J. educ. Psychol.*, 1958, **49**, 124–128.

Schaefer, E. S. A circumplex model for maternal behavior. *J. abnorm. soc. Psychol.*, 1959, **59**, 226–235.

(Received April 7, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 303-307

CONCEPT ATTAINMENT AS A FUNCTION OF NUMBER OF POSITIVE INSTANCES PRESENTED

EUGENE SHORE
Pennsylvania State University

AND

LEE SECHREST
Northwestern University

Callatine and Warren (1955) have reported an experiment in which it was demonstrated that superior performance on a concept attainment problem results from the repeated presentation of one or a few instances of the concept as opposed to single presentations of a larger number of instances. However, superior performance in this case refers to errors made during learning, i.e., associating the concept name (nonsense syllable) to the instances presented. Callantine and Warren point out that performance on *new* instances, what they refer to as transfer, was better in the group which received single presentations of 20 different instances.

The additional research evidence bearing on the problem of concept attainment and variety of instances is not totally consistent. Luborsky (1945) using an aircraft recognition task found eight exposures of a plane to be more effective than three exposures, but the five additional exposures differed qualitatively from the common three, which were diagrammatic. Reed (1946) reported more "consistent" concepts when more instances were given, but he did not control for the number of times each instance was presented to each subject. And, Adams (1954), using a problem based on spatial arrangements of lights found, somewhat suprisingly, that training on a single kind of task was better than multiple task training in a transfer test series. Morrisett and Hovland (1959) repeated Adams' study with an additional group and concluded that both a variety of problems and a high level of learning are necessary for transfer, and, presumably, concept learning to take place.

The present experiments were designed to extend the work of Callantine and Warren, who used only a single type of concept. Specifically, in the studies to be reported concepts based on nonsense syllables were chosen to represent differing levels of difficulty and differing concept attributes. The first experiment proved to involve possible methodological complications which were avoided in the second.

Experiment I

Procedure

Sixty-four male subjects from an introductory psychology class were assigned randomly to one of four experimental conditions:

Group 3 was presented with three different instances of each of four concepts, each instance being presented six times.

Group 6 was presented with six different instances of each of four concepts, each instance being presented three times.

Group 9 was presented with nine different instances of each of four concepts, each instance being presented two times.

Group 18 was presented with 18 different instances of each of four concepts, each instance being presented only once. Thus, each of the four groups was presented with a total of 72 separate stimuli, 18 for each concept. The groups differed in terms of the number of different instances of the concepts which were to be discovered.

The concepts selected for study were derived from common characteristics of certain nonsense syllables. On the basis of preliminary study, concepts at four levels of difficulty were chosen for use in the study, difficulty having been defined in terms of the inverse of the number of test instances of the concept which could be identified

after a training series on each of the concepts.

The nonsense syllables were obtained from Glaze's lists (Hilgard, 1951), with an attempt being made to utilize only syllables of low association value. Most of the instances were chosen from the 13%, 20%, and 27% association value lists. Each of the four concepts to be utilized was assigned a simple name for use in the experiment. The definitions, names, and sample instances used for each of the concepts are presented in Table 1.

The subjects were run in small groups of from 6 to 10 members each. The instances of the concepts were printed on white 4 × 6 index cards. They were exposed by hand against a neutral background at the rate of one every 13 seconds. The subjects were given 8 seconds in which to mark an appropriate answer in a booklet, were told to look up, and at 10 seconds were told the correct answer for the instance presented. The instance was exposed for 3 seconds following the reading of the correct answer. In the use of a correction procedure this experiment follows Callantine and Warren (1955).

After the presentation of the first 36 instances and again at the end of the presentation of the entire series of 72 instances, a blank card was exposed and subjects were asked to try to write the definition of each of the concepts.

Following the presentation of the series of 72 training instances of the concepts and the attempt to write the definitions of the concepts, the subjects were shown 20 additional new instances, five for each concept, and were told, "This time you will not be told the correct names of the examples. Remember, there is no penalty for guessing." The scores for each subject were the number of correct responses to each of the five test instances of the four concepts.

TABLE 1

DEFINITIONS, NAMES, AND SAMPLE INSTANCES OF CONCEPTS

Concept name	Definition	Instances
DASH	Middle letter A	KAG, ZAB, XAP, PAQ
TRIANGLE	Ending in X, Y, or Z	GOX, PIY, JEZ, KUY
DOT	Letters formed of straight lines or middle letter I or E	FIW, KEV, ZEH, TIK
CROSS	Letters can be rearranged to spell word	HES, XOP, YOJ, MUH

TABLE 2

MEAN NUMBER OF CORRECT IDENTIFICATIONS OF TEST INSTANCES OF EACH OF FOUR CONCEPTS WITHIN EACH OF FOUR GROUPS

Group	Concept				
	DASH	TRI-ANGLE	DOT	CROSS	Total
3	4.50	1.69	1.37	2.19	9.81
6	4.31	2.56	1.44	1.75	10.06
9	4.81	2.25	2.37	2.12	11.56
18	4.12	3.36	1.56	1.81	10.81
Total of means	17.75	9.81	6.81	7.87	

TABLE 3

ANALYSIS OF VARIANCE FOR REPEATED MEASURES ON TOTAL CORRECT IDENTIFICATIONS OF TEST INSTANCES

Source	*df*	*SS*	*MS*	*F*
Between subjects	63	120.44		
Between instances	3	13.50	4.50	2.52
Error (between)	60	106.94	1.78	
Within subjects	192	577.50		
Between concepts	3	204.03	68.01	35.79**
Interaction	9	30.78	3.33	1.75
Error (within)	180	342.69	1.90	

** $p < .01$.

Results

In Table 2 are presented means for the scores based on responses to the test instances. In all groups and for all concepts, identification scores exceed chance, albeit by very little in some cases. Analysis of variance (Table 3) indicated that only the difference in difficulty of the four concepts was significant. However, it became apparent that the four groups differed not only in the number of available positive instances of each concept, but in the

number of negative instances as well. Since a positive instance of DOT is also a negative instance of the other three concepts, the four groups were quite unequal with respect to negative instances. The possibility exists that an increase in negative as well as positive instances could increase cognitive strain to such an extent that information available from positive instances would be lost. A second experiment was planned to equate groups for negative instances while varying positive instances.

EXPERIMENT II

Procedure

Materials for concepts DASH, TRIANGLE, and CROSS were employed to construct new lists for three experimental conditions. In Condition 3 the subjects were exposed to three positive instances of one concept, each repeated six times, and six instances of each of the other two concepts, each instance being repeated three times. For Condition 9 subjects were exposed to nine instances of one concept, repeated twice, and six of each of the others. And, in Condition 18, subjects saw 18 instances of one concept, each only once, and again six of each of the others with three repetitions. All lists were randomized save for restrictions against repetition of an instance within a brief series.

Since the design was factorial with respect to number of instances of the concept, 3, 9, or 18, and concept difficulty, DASH, TRIANGLE, and CROSS, there were nine subgroups, each with seven subjects. The experimental procedures were much like those of Experiment I. Following presentation of all 54 instances, subjects were asked for written definitions of the concepts and were then tested with a series of eight new instances of each concept. The data for analysis are the number of correct responses for the single experimental concept during the test series.

Results

The means for number of correct identifications of the experimentally manipulated concept in each of the nine subgroups are presented in Table 4. Again the means indicate that performance was better than chance in all groups—by a somewhat greater margin than in Experiment I.

Analysis of variance which is summarized in Table 5 yielded once more a highly significant effect attributable to differences between the concepts although, as reference to Table 4 will show, the principal difference is obviously the lower difficulty of DASH. As Table 4 also indicates TRIANGLE and CROSS are reversed in difficulty from Experiment I.

Of greater interest is the significant interaction between number of instances and concept. Inspection of the means would seem to indicate that for some reason nine instances, each repeated twice are less effective for DASH,

TABLE 4

MEAN NUMBER OF CORRECT IDENTIFICATIONS OF TEST INSTANCES OF THREE CONCEPTS IN EXPERIMENT II

Condition	Concept			
	DASH	TRIANGLE	CROSS	Total
3	7.86	3.57	6.27	17.70
9	6.14	6.57	4.57	17.28
18	8.00	4.29	4.71	17.00
Total	22.00	14.43	15.55	

TABLE 5

ANALYSIS OF VARIANCE OF THE NUMBER OF CORRECT IDENTIFICATIONS OF TEST INSTANCES IN EXPERIMENT II

Source	*df*	*SS*	*MS*	*F*
Number of instances	2	0.60	0.30	0.08
Concepts	2	77.74	38.87	10.62**
Instances × Concepts	4	61.41	15.35	4.19*
Error	54	198.06	3.66	
Total	62	337.81		

* $p < .05$.
** $p < .01$.

the very easy concept, than either fewer or more instances, that the very opposite is true for TRIANGLE, and that for CROSS a few instances repeated several times are better. The results from Experiment I would support the contention that TRIANGLE is not easily solved with very few repeated instances.

Discussion

Obviously the answer to the question of optimum number and frequency of instances to be presented in a concept attainment task is still obscure. The single interaction found pointed to the probable complexity of the answer, which will very likely be shown to be a function of rather specific characteristics of the concepts involved.

In Experiment II, for example, the three concepts differed, not only in difficulty, but in the characteristics of the instances on which they were based. DASH is based on the rather obvious characteristic of nonsense syllables, the particular letters from which they are formed. TRIANGLE is based upon characteristics of the letters themselves, and CROSS requires a mental manipulation or restructuring of the stimuli presented. It may well require the repetition of an instance on several occasions in order for most subjects to eliminate the alternatives more obvious than syllables with straight letters or rearranged words.

One difficulty which still existed in the procedure employed was that the contexts were not the same for all concepts. Thus, when TRIANGLE was being experimentally treated, it was being presented in the context of a difficult and an easy concept. DASH, by contrast, was always presented in the context of two more difficult concepts. Future work should make use of a common context of instances for all experimentally treated concepts.

It is believed that the discrepancy between our results and those of Callantine and Warren (1955) may result both from the different materials used and the fact that our experiments were adapted to group procedures. Further experiments are in progress which should throw some light on questions raised above.

Summary

Two experiments were carried out in an attempt to determine whether concepts are more easily discovered from the repeated examinations of a few instances of the concept or from single examinations of larger numbers of instances. Additionally, an attempt was made to vary the difficulty level of the concepts studied. The materials used consisted of concepts based upon nonsense syllables. In the first experiment there were no significant effects attributable to the number of different instances presented, but the conditions were thought to be confounded with varying numbers of negative instances. A second experiment was performed in which number of instances and type of concept were varied in a factorial design with number of negative instances being constant across conditions. Employing recognition of test instances as a measure, a significant interaction between number of instances and specific concept was found. Although not unequivocally interpretable the results appeared to indicate that when a concept is based upon less obvious characteristics of instances or required some mental transformation of stimuli, repetition of instances is desirable. Repetition may be less necessary with relatively obvious concepts.

REFERENCES

Adams, J. A. Multiple versus single problem training in human problem solving. *J. exp. Psychol.*, 1954, **48**, 15–18.

Callantine, Mary F., & Warren, J. M.

Learning sets in human concept formation. *Psychol. Rep.*, 1955, **1**, 363–367.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.

LUBORSKY, L. F. Aircraft recognition: I. The relative efficiency of teaching procedures. *J. appl. Psychol.*, 1945, **29**, 385–398.

MORRISETT, L., JR., & HOVLAND, C. I. A comparison of three varieties of training in human problem solving. *J. exp. Psychol.*, 1959, **58**, 52–55.

REED, H. B. Factors influencing the learning and retention of concepts: III. The origin of concepts. *J. exp. Psychol.*, 1946, **36**, 166–179.

(Received May 31, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 308-316

THE ENVIRONMENTAL ASSESSMENT TECHNIQUE: A WAY TO MEASURE COLLEGE ENVIRONMENTS[1]

ALEXANDER W. ASTIN AND JOHN L. HOLLAND
National Merit Scholarship Corporation

The purpose of this report is to describe a method for measuring the college environment, and to present some data on its validity and reliability. Even though it is axiomatic that, to predict behavior efficiently, we must assess the environment as well as the person, few techniques are available for assessing human environments. Barker and Wright (1954) have described an elaborate procedure for recording the child's experiences in his community. For assessing a more restricted type of community—the college environment or press—Pace and Stern (1958) have developed the College Characteristics Index (CCI), a true-false inventory by which students can describe the characteristics of the student body, faculty, and administration of the college. More recently, Lazarsfeld (1959) and Dailey (1960) have made suggestions for research which may produce other methods to assess the environment.

Rationale

The Environmental Assessment Technique (EAT) is based on the notion, suggested by Linton (1945) and others, that a major portion of environmental forces is transmitted through other people. We can infer from this that the character of a social environment is dependent upon the nature of its members. Moreover, the *dominant* features of an environment are dependent upon the *typical* characteristics of its members. If, then, we know the character of the people in a group, we should know the climate that group creates.

As a first step, we have assumed that the college environment or "press" is a product of the following attributes of the student body: the total number of students in the college, the average intelligence of the students, and the personal characteristics of the student body (as estimated by a typology of six types). These eight variables are assumed to encompass an extensive range of the student's attributes, including his personality, interests, values, originality, self-concept, parental background, goals, and aptitudes. Assessing colleges by means of the EAT should yield environmental descriptions which will enable us to study the interactions of students and colleges in terms of a variety of educational and personal outcomes.

Environmental Measures

Explicit definitions of the eight EAT variables are given in the following sections.

Institutional Size. The size of the student body (N) is simply the total enrollment at the institution. Since the distribution of college enrollments tends to have a markedly positive skew, the square root of N was chosen as the actual measure of size. This transformation simplifies computations by reducing the number of digits and makes the distribution of college size more nearly normal.

Intelligence Level of the Student Body. To secure intelligence estimates of the student bodies, mean scores on the National Merit Scholarship Qualifying Test (NMSQT) were obtained for a sample of students entering undergraduate institutions. These students came from a 10% national sample of high school juniors scoring above the sixty-fifth percentile on the 1959 NMSQT. Only the scores of students

[1] This study is a part of the research program of the National Merit Scholarship Corporation and was supported by a grant from the National Science Foundation.

who actually entered college were used. Colleges attended by less than 15 of these students were excluded, leaving a total of 335 four-year, degree granting institutions.

The following check was made on the representativeness of differences between mean intelligence levels in the smaller samples (*Ns* between 15 and 24): mean scores on the CEEB Scholastic Aptitude Test, which were available for the entire 1956 freshman class (Fishman, 1957) at 37 of the 165 institutions enrolling less than 25 NMSQT students, were correlated with the corresponding mean NMSQT scores. The Pearson coefficient of .92 indicated satisfactory representativeness, particularly in view of the fact that different tests, given in different years, were being compared.

Personal Orientation of the Student Body. The method of personality assessment used in EAT is based on Holland's (1959b, 1961) theory of vocational choice. A central hypothesis of this theory is that persons in different occupations have different personalities. Holland has proposed a classification of occupations in terms of six personality types: Realistic, Intellectual, Social, Conventional, Enterprising, and Artistic. For each of these types there is a narrative summary of personal characteristics called the "model personal orientation." By identifying the type to which any vocation belongs, we can use a person's vocational choice as a miniature personality "test." For example, if a person's choice is engineering, which falls in the Realistic class, we would expect him to possess some of the characteristics of the model Realistic orientation: masculine, physically strong, unsociable, aggressive, etc. The validities of these personal orientations have been examined in several recent studies (Holland, 1959a, 1960, 1961).

Using this theory as a beginning, we have assumed that a given social environment can, to some degree, be described in terms of the occupations (personalities) of its members. To take a simple example: We would expect the environment or "climate" of a law firm to differ from that of an engineering firm, to the extent that we knew engineers and lawyers to be different kinds of people. If we extend this notion to the college, it should be possible to characterize the college environment in terms of the major fields (occupations) of the students.

As a first step toward characterizing the environments of the 335 institutions, each college major field was classified as belonging in one of Holland's six classes. The assignment of major fields to classes was made by using as guides the original formulations in the theory and the keys (six lists of occupational titles) in the Holland Vocational Preference Inventory (Holland, 1959a). Cases of doubt were resolved in discussion by the authors.

Descriptions of the model personal orientations and lists of the major fields included in each orientation are given in Table 1. These descriptions represent hypotheses about the climate and "press characteristics" associated with each type. Information about the number of students at a given institution who obtained a baccalaureate degree in a specific major field was secured from the publication, *Earned Degrees Conferred by Higher Educational Institutions: 1957–1958* (United States Office of Education, 1959). Unclassifiable degrees, representing about 3.2% of the total, were omitted from our analysis.

To obtain measures of the six personal orientations at any institution, the number of majors (i.e., students in a major field) of each orientation was expressed as a percentage of the total number of classifiable majors. For instance, the Realistic orientation of a college environment would be:

$$\text{Realistic orientation} = \frac{\text{Number of majors classified Realistic}}{\text{Total number of classifiable majors}} \times 100$$

In this way, each institution was characterized by the proportion of its students belonging to each of the six personal orientations, as determined by their choice of major field.

Validity of the EAT

An initial test of the EAT's validity was made using the College Characteristics Index. In the CCI, observers (usually students) at the institution describe the college in terms of 300 dichotomous items. These items are scored on 30 press scales of 10 items each. Some of these CCI data, which were obtained in 1958 from Merit Scholars and Finalists at 36 institutions,[2] have been used in two previous

[2] Amherst, Brown, California Institute of Technology, Carleton, Carnegie, University of Chicago, Colorado, Cornell, Dartmouth, Duke, Georgia Institute of Technology, Harvard, Indiana, Iowa State, Kansas, Massachusetts Institute of Technology,

TABLE 1

COLLEGE MAJOR FIELDS CORRESPONDING TO EACH OF SIX PERSONAL ORIENTATIONS

Orientation	Description (from Holland, 1961)	Relevant major fields
Realistic	"masculine, physically strong, unsociable, aggressive . . . prefers concrete to abstract"	agriculture, agricultural education, physical education, recreation, industrial arts, engineering, forestry, trade, and industry
Intellectual	"task-oriented, intraceptive, asocial, prefers to think through rather than act out; needs to understand"	architecture, biological sciences, geography, medical technology, pharmacy, mathematics, philosophy, physical sciences, anthropology
Social	"sociable, responsible, feminine . . . needs attention . . . avoids intellectual problem-solving . . . orally dependent"	health education, education of exceptional children and mentally retarded, speech correction, education (unclass.), nursing, occupational therapy, physical therapy, scholastic philosophy, social science (general), American civilization, sociology, social work
Conventional	"prefers structured numerical and verbal activities and subordinate roles . . . conforming . . . identifies with power, externals, and status"	accounting, secretarial, business and commercial (general and unclass.), business education, library science, economics
Enterprising	"verbal skills for dominating, selling, leading others . . . orally aggressive"	hotel and restaurant administration, hospital administration, history, international relations, political science, foreign service, industrial relations, public administration
Artistic	"asocial; avoids problems which are highly structured or require gross physical skills . . . intraceptive . . . need for individualistic expression"	art education, music education, English and journalism, fine and applied arts (all fields), foreign language and literature (all fields)

studies (Astin, 1961; Thistlethwaite, 1959). Each of the eight EAT measures ($\sqrt{N}$, Intelligence, and the six personal orientations) was correlated with the 30 CCI scales (Table 2). Of the 240 coefficients in Table 2, 23% are significant at the .01 level, and 39% at the .05 level.

The CCI scales most highly related to institutional size are Aggression (.64), Achievement (−.59), Counteraction-Infavoidance (−.58), Understanding (−.58), Passivity (.55), Fantasied Achievement (−.55), Deference (.54), Sex (.54), Exhibition (.53), and Pragmatism (.52). These results suggest that in the larger as opposed to the smaller student bodies there is more aggressive behavior, exhibitionism, and heterosexual activity, and more deference shown toward faculty. In smaller student bodies, on the other hand, there is more academic competi-

Michigan, Minnesota, Northwestern, Notre Dame, Oberlin, Pennsylvania, Pomona, Princeton, Purdue, Radcliffe, Rensselaer Polytechnic Institute, Rice, Smith, Stanford, Swarthmore, University of Texas, Wisconsin, Wellesley, Wesleyan, and Yale.

TABLE 2

THE RELATION OF EIGHT ENVIRONMENTAL MEASURES TO THE COLLEGE CHARACTERISTICS INDEX AT 36 INSTITUTIONS

CCI scales	Size of student body	Intelligence of student body	Personal orientations[a]					
			Realistic	Intellectual	Social	Conventional	Enterprising	Artistic
Abasement	20	−37	38	−29	04	11	−24	−45
Achievement	−59	64	−20	33	−22	−35	42	35
Adaptiveness	14	−55	14	−39	26	13	−25	−20
Affiliation-Rejection	−35	08	−21	38	−01	−07	05	16
Aggression-Blamavoidance	64	09	45	09	−25	08	−38	−58
Change-Sameness	−06	17	20	14	−14	−02	−10	−09
Conjunctivity-Disjunctivity	−19	06	−06	16	−07	−24	22	12
Counteraction-Infavoidance	−58	49	06	28	−41	−29	11	19
Deference	54	−63	26	−55	20	30	−38	−26
Dominance	−19	21	−21	32	08	−02	03	09
Ego Achievement	−02	−19	−34	−13	38	12	15	24
Emotionality-Placidity	−33	23	−25	21	−14	18	20	14
Energy-Passivity	−55	48	−05	31	−35	−42	28	33
Exhibition-Infavoidance	53	−45	−14	−33	43	36	−15	−03
Fantasied Achievement	−55	49	−32	46	−17	−21	35	38
Harmavoidance	−30	26	−50	04	21	−14	55	56
Humanism	−28	55	−81	23	25	07	79	64
Impulsion-Deliberation	27	00	11	28	−08	23	−28	−40
Narcissism	49	−36	−17	−43	59	01	01	18
Nurturance-Rejection	−03	−26	−37	−18	39	−06	22	46
Objectivity	−42	68	−26	46	−24	−26	40	31
Order	28	−51	−05	−31	36	15	−10	−03
Play	47	−60	25	−34	35	04	−36	−30
Pragmatism	52	−67	73	−32	−08	02	−73	−66
Reflectiveness	−29	50	−62	16	14	01	64	54
Scientism	−18	31	33	38	−43	−08	−22	−48
Sentience	−28	47	−70	10	31	−13	71	69
Sex-Prudery	54	−40	−22	−24	53	15	01	06
Succorance-Autonomy	−01	−28	−09	10	14	24	−14	−13
Understanding	−58	70	−28	46	−23	−24	41	33

Note.—r_{01} = .33, r_{01} = .43. Decimal points omitted in table.

[a] The percentage of college majors falling in each class.

tion, more striving for intellectual achievement, and greater involvement in campus activities (Counteraction-Infavoidance, Passivity).

Intelligence level of the student body had 18 significant correlations with the CCI scales, more than any of the other seven EAT variables. The highest correlates of intelligence were Understanding (.70), Objectivity (.68), Pragmatism (−.67), Achievement (.64), Deference (−.63), Play (−.60), Adaptiveness (−.55), Humanism (.55), Order (−.51), and Reflectiveness (.50). Most of these press characteristics are those we would expect to find in the more intelligent student bodies: an emphasis on achievement, intellectual objectivity, seriousness of purpose, and a need to analyze and understand, rather than to order and structure.

The CCI variables most closely related to the first personal orientation, Realistic, were Humanism (−.81), Pragmatism (.73), Sentience (−.70), Reflectiveness (−.62), and Harmavoidance (−.50). These correlations

imply a preference for the practical, the concrete rather than the abstract, and an aversion to feeling experience—traits which are consistent with the descriptions of the Realistic orientation (Table 1) developed earlier (Holland, 1959b).

The Intellectual orientation had its highest correlations with Deference (−.55), Fantasied Achievement (.46), Objectivity (.46), and Understanding (.46). These results imply that college environments with high Intellectual orientations have many of the expected characteristics: independence of thought and action, great need to achieve, and a need to analyze and understand.

The CCI scales correlating highest with the Social orientation were Narcissism (.59), Sex (.53), Exhibition-Infavoidance (.43), and Scientism (−.43). Here there is a close correspondence with the hypothetical description (Holland, 1961): "sociable, responsible, feminine . . . needs attention . . . avoids intellectual problem-solving . . . orally dependent" (Table 1).

The three highest correlates of the Conventional orientation, Passivity (.42), Exhibition-Infavoidance (.36), and Achievement (−.35), while of only borderline significance, are also generally consistent with the description given in Table 1.

The Enterprising and Artistic orientations are similar in the patterns of their relationships with the CCI scales: Humanism (.79 with Enterprising and .64 with Artistic), Pragmatism (−.73 and −.66), Sentience (.71 and .69), Reflectiveness (.64 and .54), Harmavoidance (.55 and .56), and Blamavoidance (.38 and .58). Perhaps one of the reasons that these two orientations turned out to be so similar, aside from the fact that they are known to be related (Holland, 1959a, p. 29) is that many of the students majoring in "business and commercial," all of whom were included in the Conventional orientation, might more properly be included in the Enterprising orientation, particularly those who plan careers in sales, promotional, and related areas.

The characteristics of the Enterprising and Artistic orientations are just the opposite of those associated with the Realistic orientation. While this is again consistent with previous descriptions (Holland, 1959a), there is the suggestion that these three orientations may be somewhat redundant.

Another partial test of EAT's validity was attempted using data reported in the book, *What College Students Think* (Goldsen, Rosenberg, Williams, & Suchman, 1960). These investigators attempted to assess and compare the values, attitudes, and opinions of students at the following institutions: Cornell, Dartmouth, Fisk, Harvard, Michigan, North Carolina, Texas, UCLA, Wayne, Wesleyan, and Yale. Even though the number of institutions was quite small, it seemed likely that the environments of these colleges would differ markedly on the variable of Intelligence. Accordingly, 16 of the questions asked of the student bodies were selected on the basis of their hypothesized relation to Intelligence. These questions dealt with: the amount of cheating admitted to by the students; preference for Republican as opposed to Democratic party; belief in God; importance of religion; attitudes toward labor laws, the welfare state, minimum wage, and a free college education; importance in college of learning values; importance of athletics; belief in the existence of academic freedom; "tendency to personify the fraternity"; and the relative importance of several possible goals of a college education—general education, vocational training, training in how to

get along with people, preparation for a happy marriage.

The institutions were ranked (1–11) on Intelligence using the sample mean NMSQT scores (*N*s ranged from 15 to 203); they were also ranked (1–11) on each of the 16 questions on the basis of the percentage of students giving positive responses, so that rank-difference correlations could be computed. Seven of these rho's were significant at the .05 level. The relevant variables and their correlations with Intelligence were:

A college ought to emphasize preparation "for a happy marriage and family life."	−.99
"College teachers are afraid to say what they really believe these days."	−.84
A college ought to "provide vocational training, develop skills and techniques directly applicable to your career."	−.81
Religion is a "major source of satisfaction."	−.80
A college ought to emphasize "a basic general education and appreciation of ideas."	.75
Republican political party preference	.73
Belief in God	−.69

On an exploratory basis, 14 more rank correlations were computed between the other seven EAT measures and two of the items which had not correlated significantly with Intelligence. Two more significant rho's were obtained: the Conventional orientation was positively related (.73) to the "tendency to personify the fraternity" and frequency of admitting cheating was negatively related (−.70) to the Enterprising orientation.

To illustrate in more concrete terms the characterization of colleges via the EAT, the three institutions (from the original sample of 335) scoring highest on each orientation were identified (Table 3). The institutions with highest Realistic orientations are technological institutions, primarily or exclusively for men; in contrast, those highest on the Artistic orientation are small liberal arts colleges for women. The highest scorers on the Social orientation include two coeducational teachers' colleges and a liberal arts college for women. Those scoring highest on the Conventional orientation are all large universities controlled by the Roman Catholic Church. (In fact, seven of the eight institutions ranking highest on Conventional are Roman Catholic institutions.)

All three of the institutions high on the Intellectual orientation are known for their strong emphasis on scholarship and research. The California Institute of Technology, despite its name, is classified as a "college, gradu-

TABLE 3

HIGHEST SCORING OF 335 INSTITUTIONS ON EACH PERSONAL ORIENTATION

Realistic
Worcester Polytechnic Institute
Michigan College of Mining and Technology
Case Institute of Technology
Intellectual
Union College and University
California Institute of Technology
Reed College
Social
State Teachers College at Towson (Maryland)
Danbury State College (Connecticut)
Georgia State College for Women
Conventional
Seton Hall University
Xavier University (Ohio)
University of Scranton
Enterprising
Georgetown University
John Carroll University
Wabash College
Artistic
College of New Rochelle
Emmanuel College (New York)
Hollins College

TABLE 4

TEST-RETEST RELIABILITIES OF THE SIX ENVIRONMENTAL PERSONAL ORIENTATIONS FOR INTERVALS OF 1, 3, AND 6 YEARS (31 INSTITUTIONS)

Orientation	Reliability		
	1 Year (1955–56)	3 Years (1955–58)	6 Years (1952–58)
Realistic	.99	.98	.97
Intellectual	.81	.55	.54
Social	.88	.87	.80
Conventional	.94	.93	.93
Enterprising	.96	.84	.83
Artistic	.95	.96	.93

ate school, and institute of research in science, engineering, and the humanities" (Irwin, 1960, p. 186).

Two of the three highest scorers on Enterprising, Georgetown University and Wabash College, are of particular interest in view of the hypothesized correlates of that orientation: "verbal skills for dominating, selling, leading others . . . orally aggressive." Georgetown University is noted for its large Law School and its School of Foreign Service. Wabash College, too, with its Personal Development Program for "young executives," its Institute of Politics, and its Washington Semester Program for "students of demonstrated ability and interest in public affairs" (Irwin, 1960, pp. 391–392), appears to place considerable emphasis on "enterprising" activities.

RELIABILITY OF THE PERSONAL ORIENTATIONS

A check on the test-retest reliability of the six measures of personal orientation was made using a 10% random sample ($N = 31$) of institutions from the original sample of 335.[3] Scores on the six personal orientations for these 31 colleges were obtained for the graduation years of 1952, 1955, 1956, and 1958 by collating the major fields listed in *Earned Degrees Conferred by Higher Educational Institutions: 1957–1958* (United States Office of Education, 1959) for those years. Table 4 shows retest reliabilities over intervals of 1, 3, and 6 years. With the possible exception of the Intellectual orientation, all measures appear to be highly reliable over the 1-year interval. Even over the 6-year period, all orientations except the Intellectual seem to have satisfactorily high reliabilities. (Table 4 also shows that differences among institutions in their curricula are highly stable from year to year.)

[3] The entire sample of colleges was not used because of the prohibitive amount of clerical labor involved in collating major fields from *Earned Degrees Conferred by Higher Educational Institutions: 1957–1958* (United States Office of Education, 1959) by college.

INTERCORRELATIONS OF EAT VARIABLES

Intercorrelations of the eight EAT variables are shown in Table 5. While all but four of the 28 correlations are significant ($p < .05$), the highest correlation is only $-.57$ (between the Realistic and Artistic orientations). An inspection of the larger correlations reveals two major clusters: the first combines Intelligence, the Intellectual orientation, and the Enterprising orientation; the second, involving the Realistic (negatively), Enterprising, and Artistic orientations, was to be expected from the patterns of CCI correlations (Table 2) discussed previously.

It should be noted that the six personal orientations are self-limiting (i.e., for any college the six scores must sum to 100). Such dependence among these measures accounts at least partially for the fact that 12 of the 15 correlations among the personal orientations were in a negative direction.

DISCUSSION

The results imply that this new method (EAT) for assessing college

TABLE 5
INTERCORRELATIONS OF EAT VARIABLES
(N = 335)

	2	3	4	5	6	7	8
1. Size	−.09	.26	−.28	−.06	.26	−.25	−.29
2. Intelligence		−.08	.41	−.36	−.27	.48	.34
3. Realistic orientation			−.24	−.37	−.17	−.46	−.57
4. Intellectual orientation				−.40	−.17	.38	.16
5. Social orientation					−.23	−.32	−.01
6. Conventional orientation						−.12	−.30
7. Enterprising orientation							.45
8. Artistic orientation							

Note.—r_{58} = .11; r_{31} = .14.

environments possesses moderate validity and substantial reliability. Moreover, EAT, being limited to eight variables which are inexpensive to obtain,[4] lessens the computational problems which arise in the study of college environments and brings such studies within the financial means of more researchers.

One limitation of the present study is that it used the major fields of the students, rather than their vocational choices. Since some of the major fields reported in *Earned Degrees* are ambiguous with regard to the actual occupations implied, it was necessary to make difficult and perhaps arbitrary decisions in assigning them to a particular class. As was noted previously, this is particularly true of those "business" majors who plan careers in sales or advertising, and should be classified as Enterprising instead of Conventional. Some of these problems would be solved if we knew the student's vocational choice.

In addition, the present scheme does not give direct information about the nature of the faculty. It is likely, however, that in level of ability, number, and "primary occupations" (fields of specialization), the faculty of an institution resembles its student body. This assumption is supported by the fact that many of the CCI items are actually descriptions of specific faculty behaviors. In addition, Pace and Stern (1958) found that faculty and student reports of the college press were similar.

EAT appears to be especially well-suited to the study of college-student interactions. A major problem in such studies is how to assess the similarity between persons and environments in a meaningful way. Since the EAT is based principally on a heuristic theory of vocational choice and occupational types, it permits us to classify both students and colleges according to the same six types. More importantly, the same *operations* (vocational choice or major field) are used for typing both the student and his environment.

The present results also indicate that the attributes of the student body reflect a major portion of what has been called the college press or environment,[5] since the press characteris-

[4] Intelligence level of the student body could, of course, be expensive to obtain; however, most institutions routinely assess their incoming freshman classes with standardized aptitude or achievement tests, and even in the absence of such test data the average high school rank of the students would probably give a useful estimate of intelligence level in the student body.

[5] It should be pointed out that such a conclusion is not necessarily at variance with studies such as those of McFee (1961), who found no evidence at one college of a systematic relationship between the needs of individual students and their perceptions of the college press. The question of the environment's being related to the nature of the students concerns *aggregates* of student characteristics *across* institutions.

tics (as measured by the CCI) of a particular school can be predicted with modest success simply from a knowledge of the students' number, average intelligence, and distribution of major fields. This conclusion is supported by a recent study of college productivity (Astin, 1961), which showed that, even for small samples of students *entering* college, the proportion planning to major in natural science was highly related to several student and faculty press variables. It is obvious that in future studies of college "influence," the students' characteristics must be taken into account before we attribute differences in performance to the college press.

One of the most compelling aspects of EAT is the possibility it offers for measuring other kinds of environments. If, as the present data suggest, a simple occupational census of the members of any group gives a valid estimate of the environment or climate of that group, it should be possible to conduct a wide variety of cultural and sociological studies using an elementary method. Studies of other institutions and cross-cultural studies, in particular, would seem especially suited to this method.

Summary

A method for measuring the college environment, the Environmental Assessment Technique (EAT), was described. Measures of eight characteristics of the student body—its size, average intelligence, and six "personal orientations"—were validated against the College Characteristics Index at 36 colleges. The six personal orientations, which were obtained from a census of students' major fields at each institution, were also shown to be highly stable over time. Some of the advantages, limitations, and possible applications of the EAT were discussed.

References

Astin, A. W. A re-examination of college productivity. *J. educ. Psychol.*, 1961, **52**, 173–178.

Barker, R. G., & Wright, H. F. *Midwest and its children*. Evanston, Ill.: Row, Peterson, 1954.

Dailey, C. A. Natural history and phenomenology. *J. indiv. Psychol.*, 1960, **16**, 36–44.

Fishman, J. A. *1957 supplement to College Board scores no. 2*. New York: College Entrance Examination Board, 1957.

Goldsen, Rose K., Rosenberg, M., Williams, R. M., Jr., & Suchman, E. A. *What college students think*. Princeton: Van Nostrand, 1960.

Holland, J. L. *Preliminary manual for the Holland Vocational Preference Inventory. (Res. ed.)* Palo Alto, Calif.: Consulting Psychologists Press, 1959. (a)

Holland, J. L. A theory of vocational choice. *J. counsel. Psychol.*, 1959, **6**, 35–45. (b)

Holland, J. L. The relation of the Vocational Preference Inventory to the Sixteen Personality Factor Questionnaire. *J. appl. Psychol.*, 1960, **44**, 291–296.

Holland, J. L. Some explorations of a theory of vocational choice. Evanston, Ill.: National Merit Scholarship Corporation, 1961. (Mimeo)

Irwin, Mary (Ed.) *American universities and colleges*. (8th ed.) Washington, D. C.: American Council on Education, 1960.

Lazarsfeld, P. F. Progress report on a project to map out the general area of nonintellectual factors in the prediction of college success. New York: College Entrance Examination Board, 1959. (Mimeo)

Linton, R. *The cultural background of personality*. New York: Century, 1945.

McFee, Anne. The relation of students' needs to their perception of the college environment. *J. educ. Psychol.*, 1961, **52**, 25–29.

Pace, C. R., & Stern, G. G. An approach to the measurement of psychological characteristics of college environments. *J. educ. Psychol.*, 1958, **49**, 269–277.

Thistlethwaite, D. L. College press and student achievement. *J. educ. Psychol.*, 1959, **50**, 183–191.

United States Office of Education. *Earned degrees conferred by higher educational institutions: 1957–1958*. (Circular No. 570) Washington, D. C.: United States Department of Health, Education, and Welfare, 1959.

(Received June 4, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 317-321

THE DESIRE FOR SECURITY: AN ELEMENT IN THE VOCATIONAL CHOICE OF COLLEGE MEN[1]

STUART H. BLUM[2]

Teachers College, Columbia University

Considerable investigation has been carried out by workers of many disciplines in attempting to determine the manner in which various factors are related to vocational choice. Factors believed to be operating in this choice are numerous and include such concepts as overall mental capacity, specific aptitudes, interests, and temperament. One of the often mentioned, but less investigated, of these factors is the desire for security.

Interviewing and counseling experience with both military personnel and civilian college students on the part of the investigator has yielded the impression that a desire for security is one of the most conspicuous factors involved in the selection of an occupation or in the choice of a specific job. Expressed emphasis by interviewees on such factors as fringe benefits, tenure, guarantee of employment, and on the use of the term security itself, often far outweighed concern about the content or nature of the actual work. Increasing trends toward a so-called Welfare State and both contemporary nonfiction and reports of scientific studies suggest an increased interest on the part of the general population in obtaining security.

The purpose of the study on which this report is based was to ascertain differences among individuals with respect to the degree of importance placed upon a desire for security as a factor in vocational choice, and to determine how this degree of emphasis is related to other pertinent characteristics or variables of the individual. Whereas any measure of security or insecurity is probably a reflection of a deeper psychological security, for purposes of the present investigation the deeper psychological aspects were largely treated as measurable normal personality variables.

The areas for investigation were based on the expectation that:

1. Degree of emphasis placed upon security as a factor in vocational choice exists as a trait of some comprehensiveness that can be measured reliably.

2. A positive relationship exists between emphasis upon security in vocational choice and the actual choice of a job situation of a secure type.

3. Significant relationships exist between emphasis upon security in vocational choice and normal personality variables as measured by the Edwards Personal Preference Schedule (1957).

4. Significant relationships exist between emphasis upon security in vocational choice and background factors.

Since an instrument to measure the degree of emphasis placed upon security in vocational choice was not available that met desired specifications, a 40-item inventory was developed.

[1] The article is based upon a PhD dissertation submitted to Teachers College, Columbia University. The author wishes to express his appreciation for the guidance and encouragement received from the members of his Dissertation Committee, Robert L. Thorndike (Chairman), Elizabeth P. Hagen, and the late Irving D. Lorge.

[2] Now at the Testing and Counseling Center, Hofstra College, Hempstead, New York.

The Security Inventory[3]

The final form of the Security Inventory contained 40 forced-choice items of the triad type. Of the three alternatives contained in each item one was scored as contributing to the Security variable which was thus expressed as a simple raw score, with a possible range from 0–40.

The content of the items was based on the investigator's interviewing and counseling experience, and utilized concepts expressed in Caplow (1954), Daugherty (1941), Hayek (1944), Hoppock (1935), Lindahl (1949), Lorge and Blau (1942), McGregor (1944), Nolan (1956), Riesman, Denny, and Glazer (1953), Roe (1956), Shartle (1952), Thorndike and Hagen (1957), Viteles (1953), and Whyte (1956). The basic conceptualization of a desire for security as expressed in the items included, when compared to other alternatives, a preference for: membership in a group; fringe and retirement benefits; freedom from the ups and downs of business or political conditions (including the concept of a guaranteed annual wage); dependence on rules, regulations, standards, procedures, and decisions of superiors (freedom from responsibility on the job); physical safety; allowance for personal shortcomings of the individual; existence of similar jobs in other organizations or under conditions of national emergency; lack of worry, risk, and competition; assurance of the maintenance of a certain standard of living; personal comfort; a fixed overall work schedule; compensation, promotion, and retention, based on amount and length of time worked; working for an organization (rather than being own boss); the proven and the established; the familiar; routine and certainty.

The booklet of the Security Inventory was entitled, *Attitudes in Vocational Choice*, and the subjects were not informed that desire for security was being measured. All items were to be completed in spite of any difficulty of making a choice.

For inclusion in the Security Inventory, the selection of the final items from an original pool of 80 was primarily based on agreement by judges as to the Security response in each item and upon a subsequent item analysis with a preliminary group of 122 male college students.

Procedure

In addition to the Security Inventory and the Edwards Personal Preference Schedule, a 24-item biographical information questionnaire which included items designed to obtain some indication of the actual job situation desired by each individual as well as his characteristics and experiences considered pertinent to the study, was prepared.

The complete set of instruments was administered to 513 students enrolled in the following institutions: City College (New York), Fairleigh Dickinson University, Colgate University, Buffalo University, and Columbia University. Of this number 404 students met the qualifications of bona fide male upperclassmen (juniors and seniors aged 30 or less). In addition to practical considerations such as availability of subjects, ease of administration of instruments, and motivation for cooperation, the male college upperclassmen group provides information and suggests implications for both education and industry as well as for psychology. Further, this group represents the potential leadership in many areas of our society at a period in time when the group's members are focusing on their vocational choices. Because the study was concerned not with absolute distribution, but rather with relationships between results on various measures, the primary concern in the selection of subjects was to secure a heterogeneous sample of upperclassmen in order

[3] A copy of the complete Security Inventory as well as a detailed description of its construction may be found in the dissertation, *Security in Vocational Choice: A Study of Male College Upperclassmen*, available from University Microfilms, Incorporated; 313 North First Street; Ann Arbor, Michigan: Order No. Mic 60-5083 remitting $3.35 for microfilm, or $4.80 for xerographic enlargement.

that adequate range of response would be obtained.

Inasmuch as the Security Inventory yielded the variable against which all the other variables examined were to be compared, the limited class time was used for the administration of this instrument. The subjects completed the Edwards Personal Preference Schedule and the biographical information questionnaire out of class on their own. A system of individual code numbers provided anonymity, and as is consistent with obtaining some degree of motivation, and with fairness and ethical standards, there was some feedback of results. Of the 404 subjects who met the qualifications for inclusion in the sample, 86%, or 346 students, completed all the materials.

Raw scores (the total number of "security alternatives" chosen) on the Security Inventory and the raw scores on the Edwards Personal Preference Schedule were correlated. Wherever possible, the items of the biographical information questionnaire were coded so as to make possible product-moment correlations with the raw scores of the Security Inventory. Analysis of variance procedures were utilized where correlational analysis was not feasible.

Results

The successful development of the Security Inventory was obviously crucial to further pursuit of the investigation. On the basis of 404 male juniors and seniors, comprising the sample of the main study before eliminations due to deficiencies on other instruments, a mean of 13.6 and a standard deviation of 6.0 was obtained for the final form of the inventory. The scores had a basically unbroken and symmetrical distribution. The reliability as estimated by Kuder-Richardson Formula Number 20 was found to be .82.

Table 1 shows the results of the analysis of variance used to test the significance of the differences among the means of the security scores for each of the categories of anticipated job situation. Progressively higher security score means are associated with progressively larger and generally more secure employers. The large size of the F ratio obtained indicates differences significant at the .01 level. It should be noted that the mean security score of those who anticipate working for themselves is approximately a full standard deviation below the mean security score of those who plan to work in civil service. The above findings also serve to validate further the Security Inventory itself, for the scores on the inventory are in the direction of the actual job situation choices.

TABLE 1

Security Score and Indication of Anticipated Job Situation

(N = 346)

Anticipated job situation	M	SD	N
Work for self	11.27	5.19	94
Work in civil service	17.94	5.80	35
Work for a large company	14.44	5.97	176
Work for a small company	13.88	5.27	41

$F = 13.07, p < .01$

Table 2 presents the correlations of the security score with each of the personality variables of the EPPS. Eight of the 15 coefficients and the multiple correlation coefficient of security score with the 15 EPPS variables are significant at the .01 level. One additional coefficient is significant at the .05 level only. Each relationship is in a direction consistent with the basic conceptualization of a desire for security. The personality characteristics most closely associated with emphasis on security in choosing a job or an occupation, are a desire for Order and an avoidance of Change.

Of the biographical variables examined by correlational analysis, two yielded coefficients with security score significant at the .01 level. High desire for security in choosing a job or an occupation shows a slight tendency to be associated with lower status of father's occupation ($r = .18$) and with frequent religious service attendance

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE EDWARDS PERSONALITY VARIABLES AND CORRELATION COEFFICIENTS, REGRESSION WEIGHTS, AND THE MULTIPLE CORRELATION COEFFICIENT OF THE SECURITY SCORE WITH THESE VARIABLES

(N = 346)

Variable	M	SD	Coefficient	Regression weight
Security	13.87	5.99		
Achievement	16.92	4.23	−.22**	−.13
Deference	10.72	3.28	.17**	.07
Order	11.05	4.44	.32**	.22
Exhibition	14.69	3.53	−.05	.06
Autonomy	14.92	4.28	−.27**	−.11
Affiliation	13.50	3.87	.09	.06
Intraception	15.86	4.69	−.06	−.02
Succorance	9.76	4.37	.21**	.11
Dominance	17.36	4.72	−.22**	−.10
Abasement	11.59	4.74	.23**	.12
Nurturance	12.82	4.58	.12*	0
Change	15.52	4.57	−.34**	−.21
Endurance	13.90	5.12	.10	.05
Heterosexuality	17.42	5.23	.02	.10
Aggression	13.97	4.43	−.08	.02
Consistency	11.45	2.09	−.05	
Multiple			.54**	

Note.—Coefficients are all product-moment.
* Significant at .05 level.
** Significant at .01 level.

TABLE 3

SECURITY SCORE AND FIELD OF MAJOR STUDY IN COLLEGE

(N = 346)

Field of major study	M	SD	N
Business	14.94	5.90	118
Engineering	13.84	4.87	49
Humanities	13.00	7.16	21
Pre-law	11.13	7.95	16
Pre-medical and pre-dental	12.12	4.23	33
Science	12.75	5.34	51
Social Studies	14.36	6.89	42
Teaching	15.81	5.57	16

$F = 2.06, p < .05$

(r = .18). Three of the variables yielded coefficients with security score significant at the .05 level only. Emphasis upon security in choosing a job or an occupation shows a slight tendency to be associated with poor health (r = .12), little athletic participation in college (r = .11), and a shorter response to the question concerning occupational goals (r = .11).

Of the biographical variables examined by analysis of variance, only the means of the security scores for each of the fields of major study in college gave an F ratio significant at the .05 level.

The hypothesized expectations were thus confirmed, and for college upperclassmen at least, it can be concluded that emphasis upon a desire for security in choosing a job or an occupation exists as a trait of some comprehensiveness that can be measured reliably, is positively related to the actual choice of a job situation of a secure type, and is characterized by low, but statistically significant personality and background correlates.

DISCUSSION

It is likely that what might be classed as chance or accident plays an important role in an individual's selection of a job or an occupation. Nonetheless, previous investigations have pointed to the relative importance of the desire for security as a factor in this decision. The investigation on which this report is based has demonstrated that although statistically significant relationships exist between the variable and both normal personality variables as measured by the EPPS and background factors, these relationships are small. For practical purposes, therefore, it would often seem desirable to include a measure of an individual's desire for security in choosing a job or an occupation as a part of vocational guidance and personnel selection procedure. The results of a more comprehensive and clinically derived measure of basic phychological security-insecurity might well possess a high positive relationship with the score obtained on the Security Inventory. However, from the view-

point of administrative efficiency, as well as from that of validity for the vocational guidance and/or personnel selection situation, the use of the more specific device appears indicated.

The differences among security score means for each of the fields of major study in college provide the basis for an additional measure in guidance practice. For example, a high security score would be deviant for a pre-law student, and might suggest further interview, since the group comprising this academic major possesses the lowest mean score.

It may be argued that the measure of desire for security is contaminated by the educational or occupational choice the person has already made. When an individual has chosen a field of major study in college that prepares for a specific profession or vocation, he has to some extent made an occupational commitment. He may have chosen the field for reasons other than a desire for security. Yet the choice may have influenced him to select alternatives that were consistent with this choice when filling out the Security Inventory. The individual who makes an occupational commitment, presumably on the basis of other than security reasons or motives, nonetheless, has probably in some manner taken security into account. It has not been determined which of these two elements of motivation takes precedence.

Summary

One of the often emphasized, but less investigated factors related to vocational choice is the desire for security. A 40-item inventory designed specifically to measure the degree of emphasis placed upon security in vocational choice by male college upperclassmen, was developed. Results on this Security Inventory, the Edwards Personal Preference Schedule, and a questionnaire to ascertain background information were obtained from a heterogeneous group of 346 upperclassmen. It was found that emphasis upon a desire for security in choosing a job or an occupation, exists as a trait of some comprehensiveness that can be measured reliably, is positively related to the actual choice of a job situation of a secure type, and is characterized by low, but statistically significant, personality and background correlates.

REFERENCES

Caplow, T. *The sociology of work.* Minneapolis: Univer. Minnesota Press, 1954.

Daugherty, C. R. *Labor problems in American industry.* (5th ed.) New York: Houghton-Mifflin, 1941.

Edwards, A. L. *Manual of the Edwards Personal Preference Schedule.* (Rev. ed.) New York: Psychological Corporation, 1957.

Hayek, F. A. *The road to serfdom.* Chicago: Univer. Chicago Press, 1944.

Hoppock, R. *Job satisfaction.* New York: Harper, 1935.

Lindahl, L. G. What makes a good job? *Personnel,* 1949, **25,** 263–266.

Lorge, I., & Blau, R. D. Do the better qualified compete for public jobs? *Publ. personnel Quart.,* 1942, **3,** 29–36.

McGregor, D. Conditions of effective leadership in the industrial organization. *J. consult. Psychol.,* 1944, **8,** 55–63.

Nolan, C. Y. Attitude differences among disparate Air Force specialties. *USAF Personnel Train. Cent. res. Rep.,* 1956, No. TN-56-88.

Riesman, D., Denny, R., & Glazer, N. *The lonely crowd: A study of the changing American character.* (Abr. ed.) Garden City, N. Y.: Doubleday, 1953.

Roe, Anne. *The psychology of occupations.* New York: Wiley, 1956.

Shartle, C. L. *Occupational information: Its development and application.* (2nd ed.) New York: Prentice-Hall, 1952.

Thorndike, R. L., & Hagen, E. P. Attitudes, educational programs, and job experiences of airmen who did not reenlist. *USAF Personnel Train. Cent. res. Rep.,* 1957, No. TR-57-2.

Viteles, M. S. *Motivation and morale in industry.* New York: Norton, 1953.

Whyte, W. H. *The organization man.* New York: Simon & Schuster, 1956.

(Received June 21, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 322-329

THE EFFECTS OF MANNER OF SPEECH ON APPRECIATION OF SPOKEN LITERATURE

GEORGE M. GLASGOW[1]
Fordham University

Esthetic appreciation is recognized as an important objective of education and condition of artistic development. In the study and teaching of literature, there is a need for identification and measurement of elements of communication which promote the appreciation of literature. Hence, the present investigation has been designed to measure the effects of certain speech characteristics on audience appreciation of descriptive prose and lyric poetry. The problem is to determine whether or not variations in the manner of speaking the literature increase or decrease audience appreciation.

Theoretically, the problem has assumed psychological relevance with the appearance of a thoughtful survey of the literature on esthetics by Pratt (1961). He refers to the concept of "tertiary qualities" (p. 80): "the expression of art as a property closely bound with the perceptual structure" of the person who responds as an aspect of esthetic theory to be explored further. In this inquiry, the expressive qualities of elements of speech have been varied in terms of postulated "good" or desirable and "poor" or undesirable modes of presentation. The operations for exploring the esthetic experiences of human observers are the elicitation of discriminatory reactions upon hearing the varied presentations.

[1] The author wishes to express gratitude for the review of this manuscript by Dorothea McCarthy.

Problem

The expressive elements of speech tested were voice quality, pitch, rate, and enunciation. Within each of these elements the effects of a good manner were contrasted with those of a poor manner in terms of their influence on audience appreciation. Each speech element was treated as the only variable in one experimental unit. There were thus four experimental units. In the first unit, the effects of a good voice quality were compared with the effects of a nasal twang quality. In the second unit, varying pitch was compared with a monopitch. The third unit treated moderate rate versus fast rate; and the fourth unit, distinct enunciation versus indistinct enunciation.

A difference exists between the direct color stimuli of a painting, or the audible tones of music, and imagined color or audible stimuli which result from hearing spoken-word image symbols. There is but one hue or shade of red at any point of a picture. But the theoretical number of specific hues or shades of red which may be imagined in response to the spoken word "red" may be many and may be limited only by the spectra of one's imagination and the cumulative layers of experience. The present investigation explores the relationship between the sound spectra of voice and speech and associated variations in audiences' visual and auditory images, moods, ideas, and literary values.

Appreciation was considered as a

mental and emotional response to an object, situation, or condition in which percepts, concepts, images, and evaluations are personalized in the meanings and feelings of the appreciator. Appreciation has been viewed as self-expression. "The lyric poem is appreciated by us as an expression of our inner life" (Parker, 1920, p. 40). As self-expression, appreciation would seem to overlap "creativity," explained by Rhodes (1961, p. 305) as a "phenomenon by which a person communicates a new concept (which is the product)."

The process of appreciation may be treated analytically and synthetically. Parker (1920, pp. 188–193) has referred to the elements of literary appreciation, e.g., images, sound sensations, meanings, feelings, the rhythm of ideas; but these find embodiment and unity in the meaning of the whole esthetic experience.

The measurement of appreciation and esthetic experience has been studied by numerous investigators. Speer (1929) conducted research on the "Measurement of Appreciation in Poetry, Prose, and Art and Studies in Appreciation." Langfeld (1920) and Birkhoff (1933) evaluated the "aesthetic attitude" and esthetic measures. Recently Thurstone (1959) described a method of measurement which seems to be basically comparable to one of the methods used in the present study, as follows:

> Some time ago I obtained some data to ascertain whether we can make an experimental determination of the zero point in a scale of preferences. The zero point would be that point at which the subject is neutral as to whether he wants the object or not. Any object in the scale below the neutral point is then an object which the subject dislikes.... We have made an initial analysis ...and the results are very satisfactory (p. 181).

In this investigation, good quality was a pleasant, male baritone quality. The nasal twang was produced by a constriction of the nasopharynx. Good pitch was an appropriately varying intonation. The monopitch was a sustained pitch of a piano "middle C." The good rate was 130 words per minute; the fast rate was 250 words per minute. Good enunciation was distinct but not "artificial" or overprecise; poor enunciation was moderately indistinct and portions of it were slightly slurred.

Method

The investigation consisted of two one-group experiments conducted with two similar but different groups. The second experiment was essentially a replication of the first, with the exception of changes in the order of presenting the good-poor sequences and their specimens. The purpose of repeating the experiment with a second group was to secure additional evidence and to provide opportunities for changes in the order of rotation.

In each experiment, the "good" level of a speech element, e.g., quality, was applied to one member of a pair of literary specimens and presented to an audience and their appreciation of the spoken specimen was measured. Then the "poor" level of the same speech element was applied to the paired mate of the first specimen and presented to the same audience and their appreciation of the second specimen was likewise measured. If the good level of a variable was presented first, the poor level immediately followed, and if the poor level was presented first, the good level immediately followed.

When the good level of any variable was presented in a "set," all other variables and all other recognizable speech characteristics in the same good presentation were kept good as far as was possible. When the poor level of the same variable was presented, all speech factors remained good as far as possible but the single, poor speech level. Thus, any differences between the audience's appreciation of the first and the second specimens spoken in the good and the poor speech levels were attributed to the different effects of these levels on appreciation.

Literary Specimens. It was recognized

that the two literary specimens used to express the two different levels of any variable should be alike in order that the same potentials of appreciation would be available on which these good and poor levels could exert their possibly varying expressive influences. Exact duplicates, however, would have entailed the use of the same literary specimen for any good-poor set, since every literary specimen is unique, and this would have involved the repetition of "content" for every second presentation in a good-poor set. Consequently, the use of a pair of similar matched specimens was considered to be preferable, conditioned with certain safeguards to enhance the value of the results.

First, two pairs of prose and poetry specimens were used for each speech element instead of the minimum of one pair. Second, a rotation was employed for the order of presentation of the good and poor levels of each element, for the speech elements themselves, and for the particular specimen in a pair used to express the good or the poor level of a speech element. It may be noted that "The rotation method ... is frequently used when ... there is doubt concerning the equivalence of the groups" (Good, Barr, & Scates, 1938, p. 495). In this experiment, there was doubt concerning the equivalence of the two specimens of each literary pair. Hence the three factors mentioned were rotated. Because of the nature of the experiment, it was not necessary to equate the two groups, subject against subject. Third, although the experiment with the first group was complete, the basic experiment was repeated, as indicated, in order to secure additional evidence and to provide opportunity for further rotation of the speech levels and specimens.

Pairing of Specimens. The matching of the literary specimens in each pair[3] was done by 10 experienced high school and college English teachers on the following bases: suitability of content for the grade level, similarity of literary style, potentials for appreciation, length, literary value, and relative unfamiliarity to the subjects. The pairs were designated: A-B, C-D for quality; E-F, G-H for pitch; I-J, K-L for rate; and M-N, O-P for enunciation.

Subjects. The two groups of subjects consisted of 90 and 136 fourth term high school sophomore girls. The groups approximated each other to the following extent: All members of both groups were of the same grade level and sex. Only those with normal hearing, as determined by two audiometer tests, were included in either group. The mean IQ for Group I was 105.5 and for Group II, 105.1, with standard deviations of 11.9 and 10.3, respectively. There were no significant differences in the distributions of the IQ scores. English grades were also considered as possible indexes of the students' potential to appreciate English literature. Averages of four English grades yielded a mean of 78% with a range of 67–91% for Group I and 76% with a range of 63–91% for Group II. In view of the mean IQ and mean English marks, it was believed the two groups were similar samples of female high school students for purposes of this experiment.

Judging of Specimens. Three experienced university speech teachers agreed unanimously as to the goodness and poorness of the opposite levels of each variable. They also agreed as to the basic goodness and sameness of the other recognizable speech aspects in both presentations of each "set" for each variable. The good characteristics were judged to be esthetically pleasing, to express suitable emotional states, to sustain attention, and to provide appropriate intonation, emphasis, etc. The good manners were believed to be better than the average comparable characteristics of college students; the poor manners moderately worse.

Recordings. High-fidelity recordings were used in order to keep the presentations constant, to insure that the speech examples which were approved were the ones used during the procedure, and to isolate speech as the experimental factor apart from gestures and facial expression. All the speech examples were recorded by the writer.

Rotational Sequence. The pairs of literary specimens were presented in good-poor sets. A set was a pair of specimens with one member spoken in the good level and the other in the poor level of the same speech element. The sets were presented to Group I in two series as follows: quality, A good-B poor; pitch, G poor-H good; rate, I good-J poor; enunciation, O poor-P good. In the second series the second set of specimens for each variable was presented in the opposite set sequence, namely, quality, C poor-D good; pitch, E good-H poor; rate, K poor-L good; enunciation, M good-N poor. The sequence was the same for prose.

In the procedure with Group II, the set sequence was the reverse of that used for Group I. Specimen A was presented with poor quality and B with good quality; G with good pitch modulation and H with

[3] Typical of a "pair" of poetry specimens were "Love Song from New England" by Winifred Wells and "I Shall Not be Afraid" by Aline Kilmer. Typical of the prose specimens were selections from Anne Lindbergh's *North to the Orient*. Prose specimens were limited to 260 words each.

monopitch. The remaining variables followed a similar order.

Measurement of Appreciation. The experiment required an appreciation test which would show changes of expressed appreciation in accordance with differences in the good and poor presentations of the variables. Since such tests were not available, it was necessary to construct original tests for this study.

The tests were basically check sheets having steps representing different degrees of response. Two techniques of measuring appreciation were included in each test given. The first technique secured as indexes of appreciation the listeners' estimates of one of three degrees of the *presence* of each of five specified appreciation elements in each selection heard *plus* the extent to which the listeners *liked* each element. The three degrees of presence from which one was to be checked were strong, medium, and weak. The three steps of liking for the elements were strong, indifferent (neutral), and disliking.

The appreciation elements listed for prose were recognition of topic, clarity of thought, unity, vitality, and sense imagery. Those for poetry were recognition of theme, mood, force of emotion, sense imagery, and rhythm.

The second technique of testing secured the listeners' estimates of the *literary merit* of each whole specimen heard *plus* their personal preference for each whole specimen. Five levels of literary merit were mentioned: very superior (a masterpiece), superior, average textbook material, standards of a less important magazine, and below newspaper standards. In treating the data later the first two levels were combined into one termed "high" and the last two into one called "low." Hence, literary merit as treated became high, medium, and low. The levels of preference were also high, medium, and low.

Results

Appreciation Elements

The first step in treating the data was the combination of each score on the "presence" of each appreciation element with the parallel score on liking in order to form a joint index of "perception" plus emotional response for each specimen. This provided five indexes for every poetry and prose selection heard by each subject.

Since each presence score could be combined with any of the three degrees of liking, there were nine possible combinations of presence and liking for each of the five elements in every selection. These were: strong presence, Number 1, with any of the three degrees of liking; medium presence, Number 2, with any of the three degrees of liking; and weak presence, Number 3, also with any of the three degrees of liking. In such a circumstance, the presence-liking ratio provided three cells of perfect agreement, namely 1:1, *high*; 2:2, *medium*; and 3:3, *low* appreciation.

The scores for the two good specimens in each two sets for each variable were averaged and the scores for each two poor specimens opposed were likewise averaged to form good and poor categories. The percentages of the scores in each of the nine presence-liking cells were computed for the good and the poor manners of each variable in prose and in poetry. These percentages showed that from 75–87% of all scores had been placed in the cells of perfect agreement: the 1:1, the 2:2, or the 3:3.

The data showed that the percentages of subjects expressing high appreciation of the individual elements were consistently greater when the speech manner was good than when it was poor. The number expressing low negative appreciation was consistently greater when the speech manner was poor.

The next step was the computation of an appreciation index which was an average of all five elements for each specimen, good or poor, heard by each subject.

These indexes for both groups of subjects were so similar that the scores of both groups in the same categories were combined from here on. The distributions of average response scores of Group I and Group II were found

TABLE 1

PERCENTAGES OF SUBJECTS IN THE HIGH, MEDIUM, AND LOW APPRECIATION CELLS BY TYPE OF SPEECH PRESENTATION OF EACH VARIABLE IN PROSE AND POETRY AND THE CHI SQUARES

Appreciation	Quality		Pitch		Rate		Enunciation	
	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Prose								
High positive	40	19	44	27	55	29	47	5
Mediocre	33	38	30	32	23	33	31	29
Low negative	4	17	5	15	3	15	3	53
Chi squares*	66.61		37.23		69.11		337.65	
Poetry								
High positive	41	18	42	16	44	15	44	8
Mediocre	26	34	29	29	31	35	33	32
Low negative	4	17	5	27	3	25	2	36
Chi squares*	78.15		114.32		135.85		231.78	

* All chi squares are significant at beyond the .001 level of confidence.

to conform very closely to each other in comparable categories. Eight comparisons were made, namely, four for the presence scores and four for the preference scores. The four comparisons of presence-score distribution were composed of two prose comparisons, good and poor, and of two poetry comparisons, good and poor. The same comparisons were made for preference. In none of the eight instances, could the "null hypothesis" of "no difference" be rejected at the 5% level of confidence. The comparisons were made by means of the Kolmogorov-Smirnov two-sample test (Siegel, 1956, p. 131). No differences between the cumulative percentage distributions of Groups I and II exceeded the critical value of 18%. These findings were considered sufficient for assuming that the scores for Groups I and II were drawn from the same population and could therefore be combined.

Table 1 lists the percentages of the total number of subjects who expressed high positive (1:1), mediocre (2:2), and low negative (3:3) appreciation in each good and poor category of each variable and the chi squares for the "independence" of the good and the poor speech versus the three levels of appreciation.

In Table 1 variations in appreciation scores accompanying variations in speech manner are evident. The number of subjects who expressed high appreciation was markedly and consistently greater when the speech manners were good than when they were poor. In prose, for example, the good-poor percentage ratios of 40:19 for high appreciation of quality through to 47:5 for enunciation reveal two to nine times as much high appreciation when the speech styles were good as when they were poor. Similar ratios exist in poetry. Moreover, low negative appreciation of prose and poetry was greater when the speech manners were poor.

There is an interesting similarity between the parallel scores for prose and poetry. For instance, the good-poor ratio for high appreciation under quality is 40:19 for prose and 41:18 for poetry; enunciation, 47:5 for prose and 44:8 for poetry. Such percentages

would seem to indicate that speech manners affect appreciation somewhat similarly even though the types of literature may differ.

Whole Literary Specimens

The percentages of the subjects' estimates of the literary merit of, and preferences for, the whole specimens are presented in Table 2. Since the high and low levels of merit and preference tell the essential story, the middle level is omitted here. Chi squares are listed for the independence of the high, medium, and low levels of literary merit and preference, respectively, versus the good and the poor styles. Chi squares were also computed for the good and the poor frequency deviations of the combined four variables from their theoretically equal (1:1 basis) distribution on the high appreciation and the low appreciation levels. The chi square for high appreciation was 2000.52 and for low appreciation, 2862.84. As the chi squares indicate, the differences are highly reliable.

Table 2 shows the reduction of the subjects' ratings of the literary merit and their personal preferences for the specimens occurring concomitantly with the poor style of each speech variable. Nasality, monopitch, fast rate, and indistinct enunciation notably *decreased* the number of persons who rated the literary merit high in comparison with the number when the speech manners were good. The number of ratings of low literary merit and of low preference for specimens was consistently greater when a poor

TABLE 2

PERCENTAGES OF SUBJECTS ASSIGNING HIGH AND LOW LITERARY MERIT TO LITERARY SPECIMENS AND PREFERENCES FOR THE SPECIMENS IN EACH MANNER OF SPEECH PRESENTATION AND THE CHI SQUARES

	Quality		Pitch		Rate		Enunciation	
	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Prose								
Merit:								
High	20	6	19	7	32	13	25	4
Low	31	56	36	57	20	41	29	81
Chi squares	67.76		45.89		61.14		235.14	
Preference:								
High	30	11	30	16	48	22	39	5
Low	10	35	11	27	7	23	7	65
Chi squares	97.40		44.58		76.29		333.00	
Poetry								
Merit:								
High	40	14	34	9	31	11	53	12
Low	24	52	22	56	25	53	13	63
Chi squares	97.26		126.77		86.20		249.80	
Preference:								
High	45	18	49	11	47	19	63	13
Low	7	32	4	29	3	30	1	45
Chi squares	116.41		186.52		143.33		315.77	

speech level was used. This held true for all four speech variables. Poor enunciation was associated with greater changes in appreciation than any other variable.

Biserial Correlations

As evidence concerning the significance of the deviations of the good versus the poor levels of each speech element, biserial correlations were computed to show the extent of the differentiation between the two levels as they operate through quality, pitch, rate, and enunciation of both prose and poetry. The biserial correlation coefficient has been interpreted as a measure of the degree of divergence of the effects of the good and the poor levels from equality.

In computing these coefficients, each subject's average of the two good scores for the presence of the appreciation elements was combined with the average of the two good literary merit scores rated for each variable in order to form a composite presence score. The same was done with the scores in the poor category. Similarly, each subject's scores for preference for the appreciation elements were combined with her preferences for the whole selections in order to form a single preference score for the good and the poor categories, respectively, for each variable. These coefficients appear in Table 3. As may be seen, the coefficients for enunciation are considerably higher than for any other speech element, a finding which suggests the greater effect of enunciation upon appreciation.

TABLE 3

THE BISERIAL CORRELATION COEFFICIENTS FOR THE GOOD AND THE POOR LEVELS FOR PRESENCE AND PREFERENCE IN PROSE AND POETRY

Variable	Presence		Preference	
	Prose	Poetry	Prose	Poetry
Quality	.55	.70	.55	.71
Pitch	.46	.73	.44	.78
Rate	.53	.71	.55	.76
Enunciation	.95	.94	.95	.93

SUMMARY

A "good" and a "poor" manner of each of four speech variables (voice quality, pitch, rate, and enunciation) were used to present 16 paired specimens of prose and 16 specimens of poetry to 226 high school sophomores. The speech manners were good quality versus nasal twang, varied pitch versus monopitch, moderate rate versus fast rate, and distinct versus indistinct enunciation. A markedly greater number of subjects expressed high appreciation of and strong preference for the selections, and also rated the selections as of high merit, when the speech manners were good than when they were poor.

The number of subjects who rated the literary merit high was greater under the good speech manner than under the comparable poor manner of every variable. The number who rated the literary merit low was greater under every poor speech manner.

The number of subjects expressing strong preferences for the whole selections was increased by every good speech manner and decreased by every poor speech manner.

Poor enunciation decreased high appreciation and increased low appreciation in terms of the comparable good manner, more than the poor manner of any other speech element.

All the chi squares are significant at beyond the .001 level of confidence.

Thus, different manners of speaking literature have different effects on audience appreciation and audience judgments of literary merit. Consequently, it is believed that speech manner is an

important factor in the educational development of potential literary appreciations and insights.

REFERENCES

BIRKHOFF, G. D. *Aesthetic measure.* Cambridge: Harvard Univer. Press, 1933.

GOOD, C. V., BARR, A. S., & SCATES, D. E. *The methodology of educational research.* New York: Appleton-Century, 1938.

LANGFELD, H. *Aesthetic attitude.* New York: Harcourt, Brace, 1920.

PARKER, D. H. *The principles of aesthetics.* New York: Silver Burdett, 1920.

PRATT, C. C. Aesthetics. *Annu. Rev. Psychol.*, 1961, **12**, 71–92.

RHODES, M. An analysis of creativity. *Phi Delta Kappan*, 1961, **42**(7), 305–310.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SPEER, R. Measurement of appreciation in poetry, prose and art and studies in appreciation. *Teachers Coll. Contrib. Educ.*, 1929, No. 362.

THURSTONE, L. L. *The measurement of values.* Chicago: Univer. Chicago Press, 1959.

(Received June 21, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 6, 330-334

MEASURING THE VALUE SYSTEMS OF EDUCATION PROFESSORS[1]

DONALD J. VELDMAN, ROBERT F. PECK, AND CARSON McGUIRE
University of Texas

The basic idea of this study can be applied to any situation where two or more judges make quantified assessments of people on a series of characteristics. Most factorial studies analyze the relationships between scales (*R* technique) or between subjects (*P* technique) after averaging or combining the scores given by different judges. The point of the present study, however is this: If each judge's ratings on the total population are separately factor analyzed, it becomes possible to identify his particular way of construing the characteristics under study. Furthermore, his assumptions (often unconscious) about the way he thinks the characteristics relate to one another are directly shown by the factors that emerge from his ratings. The present method resembles Kelly's (1955) "personal construct" approach, but is applicable to *any* set of judgments, regardless of their specific content.

If, upon comparison, all or most of the judges turn out the same factor structure, they may be said to show high reliability. Moreover, such agreement can be regarded as evidence of construct validity for the dimensions of human nature on which they all agree. It must be remembered, however, that agreement does not necessarily prove the common set of conceptions to be valid. All it reflects is a common frame of reference among the judges. Validation against external evidence does continue to have a persuasive desirability of its own.

A very interesting possibility arises, however, when *dis*agreement is uncovered. When comparison of the several factor-systems reveals significant differences between the individual judges, it is possible to interpret each judge's "unique" factors as a revealing projection of his own special, somewhat private conceptual world. Indeed, it is possible to identify precisely how many of his concepts he holds in common with his colleagues, and precisely how much he differs, where he does differ. Moreover, the *pattern* of relationships the judge expects to obtain among the different aspects of human nature is clearly specified. Whether any given judge is "right" or not, is a subject for further determination; but what each judge currently *believes* to be true is made clear.

When the measured characteristics include many aspects of personality which presumably are related to teaching effectiveness, and the judges are professors of education, assessing prospective teachers, the findings take on considerable practical significance (Peck, 1960). At least, what professors *believe* to be true about human nature is apt to influence their perception and evaluation of their students. What they believe is also likely to color their teaching and their advice to school executives.

Finally, when the measures include an estimate of the "overall mental health" of each subject, it becomes possible to determine exactly what "mental health" means to each judge, in terms of all the behavioral characteristics included in the study.

[1] This research is a part of the Mental Health in Teacher Education Project, College of Education, University of Texas. This project is supported by Grant No. 2M-6635 from the Training Branch, National Institute of Mental Health.

Procedure

Rater Constructs. Four judges experienced in personality assessment (two professors, two research associates) independently rated 53 female sophomores in education. Fifty aspects of personality and behavior were rated on seven-point scales. A rating on "overall mental health" was included.

The appraisals were based on the following battery:

(1) Biographical Information Form
(2) Peck Sentence Completion, 2-D (90 items)
(3) Group Thematic Apperception Test (8 pictures)
(4) Bown School Situations Analyzer
(5) Educational Issues Questionnaire

Following completion of the assessments, the ratings were transferred to IBM cards and all subsequent computations were carried out with an IBM 650 data processing system.

The 50 rating scales were intercorrelated for each rater separately. The four resulting intercorrelation matrices were separately subjected to centroid factor analysis and the resulting structures were rotated by means of the normalized varimax method which maintains orthogonality among the axes.

The factors extracted in each analysis are described in Table 1, which also includes the percentage of common variance accounted for by each factor, and the loading of the overall mental health rating on each factor.

In order to obtain an objective estimate of the degree of similarity among the various rater factors, the rotated factor loadings were used as data for an intercorrelation analysis of the 25 factor variables. Table 2 indicates the primary interrater factor relationships.

Defining Mental Health. In Table 3 the major personality correlates of mental health are listed. These constitute the common area of agreement about the meaning of "mental health." Table 4 shows the personality variables on whose relationships to mental health the four judges differed.

TABLE 1

Factors Emerging in Each Analysis, the Common Variance Associated with Each, and Their Mental Health Significance

Factor	Factor description	Percentage of common variance	Mental health loading
Rater A			
I	Creative, perceptive, independent	29	.58
II	Emotionally and judgmentally restrained	17	.43
III	Optimistic, relaxed, peer oriented, sociable	24	.59
IV	Child oriented, affectionate, indecisive	10	.12
V	Sexually integrated	2	.18
Rater B			
I	Intelligent, perceptive, independent, creative	30	.67
II	Judgmentally and emotionally restrained	15	.29
III	Sociable, peer oriented	9	.37
IV	Child oriented, affectionate, cheerful	11	.40
V	Self-confident, hopeful, relaxed	5	.27
VI	Nondogmatic, flexible	3	.13
VII	Conventional, stable, expedient	2	.15
Rater C			
I	Independent, dominant, nonconforming	33	.74
II	Judgmentally and emotionally restrained	13	.41
III	Affectionate, cooperative, peer oriented	19	.34
IV	Mild emotional reactivity	3	.02
V	Imaginative, insightful	2	−.05
VI	Goal oriented, emotionally constricted	3	.01
VII	Hopeful, sociable, rigid, "adult"	2	.04
Rater D			
I	Creative, independent, perceptive	27	.63
II	Stable, emotionally and judgmentally restrained	16	.35
III	Sociable, cooperative, affectionate, cheerful, peer oriented	20	.54
IV	Persistent, proud, "adult"	5	−.02
V	Relaxed, self-confident	3	.06
VI	Rational, sexually integrated, goal oriented	4	.22

Results

Rater Constructs

It is apparent from Table 1 that the judges employ construct systems which vary considerably in degree of differentiation. (The percentages of common variance are a more reliable index of meaningful differentiation than is the number of factors extracted in each analysis.)

Table 2 reveals the presence of three dimensions common to the construct systems of all four raters:

TABLE 2

PRIMARY INTERRATER FACTOR RELATIONSHIPS

Factor	Rater A	Rater B	Rater C	Rater D
I	B_1 97	A_1 97	B_1 94	C_1 94
II	B_2 92	A_2 92	A_2 90	A_2 92
III	C_3, D_3 86	A_3 86	D_3 91	C_3 91
IV	B_4 69	C_3 87	D_2 42	C_7 55
V	C_1 47	A_3 78	B_1 43	B_5 70
VI	—	C_3 66	D_5 33	C_1 57
VII	—	D_4 37	D_4 55	—

Note.—This table is a condensation of a square matrix of factor loading intercorrelations, and includes only the highest coefficients in each column of the matrix. Decimal points have been omitted.

Common Dimension I: Creatively Intelligent Autonomy. Factor I accounted for the most variance in all analyses. Comparisons of the descriptions in Table 1 reveal only minor differences in emphasis among the four raters, except for the absence of intellectual attributes in Rater C's cluster.

Common Dimension II: Emotional-Judgmental Restraint. The key traits here are virtually identical for all raters, although Rater D placed somewhat more weight on stability of mood.

Common Dimension III: Peer Oriented Sociability. All four raters agree on a core cluster of traits, with wide variations in the other variables they also include in this dimension. Rater B appears to define the dimension most narrowly, while Rater D allows a large number of traits nearly equal weight.

Unique Dimensions. Beginning with the fourth factor in each analysis, the dimensions become more idiosyncratic. In some cases it is possible to identify a separate dimension for one rater as part of a broader dimension in another rater's system. Considering the coefficients in Table 2, on the other hand, it would be difficult to regard factors A_5, B_7, C_4, C_5, C_6, D_4, or D_6 as anything but idiosyncratic.

Factors A_4 and B_4 are quite similar and seem to represent as a separate cluster the characteristics of affection and positive orientation toward children which Raters C and D include in their third common factor. Factors B_5 and D_5 appear to be a good deal alike. They separate off one facet of Rater A's third factor—relaxation—which only he includes in Factor III. Factor B_6, although somewhat unique to this rater, represents "flexibility," a major

TABLE 3

PRIMARY CORRELATES OF MENTAL HEALTH, AVERAGED OVER FOUR RATERS

Average correlation with mental health scale	Bipolar scale description		
82	Rational-altruistic character	vs.	Nonrational-altruistic character
82	Objective, accurate perception	vs.	Distorted, biased perception
81	Rational, objective judgment	vs.	Biased, personalistic judgment
78	Orderly, rational thought	vs.	Confused thinking
77	Basic trust	vs.	Distrust of people and life
76	Full reality awareness	vs.	Constricted undifferentiated perception
75	Self-insightful	vs.	Little self-insight
74	Appropriate goals	vs.	Inappropriate goals
73	High self-esteem and self-confidence	vs.	Low self-esteem, self-doubting
73	Appropriate emotional expression	vs.	Inappropriate emotional expression
72	Relaxed, comfortable	vs.	Anxious, many specific fears
71	Positive regard for family	vs.	Negative attitude toward family

TABLE 4

SCALES OF UNCERTAIN SIGNIFICANCE FOR MENTAL HEALTH

Correlation with mental health scale*				Bipolar scale description		
A	B	C	D			
28	56	48	10	Rich, complex fantasy life	vs.	Dull, unimaginative
54	77	54	18	Decisive, definite judgment	vs.	Ambivalent, tentative judgment
38	58	09	41	Affectionate, kind	vs.	Undemonstrative of affection
74	53	67	82	Genuine interest in others	vs.	Self-centered
53	60	62	17	Dominant, ascendant, competitive	vs.	Submissive, following, yielding
83	81	59	73	Appropriate goals	vs.	Inappropriate goals
58	74	80	47	Persistent, tenacious	vs.	Quitting, gives up easily
11	54	03	−13	Strong identification with children	vs.	Little identification with children

* Highest and lowest coefficients differ significantly at the .01 level.

aspect of the third dimension for Raters C and A.

Although the idiosyncratic dimensions which are unique to single raters will not be discussed here, such dimensions are of special value in identifying conceptions which the individual does not share with anyone else. This information could be quite useful, for instance, in training assessment specialists to identify and correct conceptual sets which are mainly projections and which may distort their judgment.

Mental Health Definition

Table 3 shows the way the term "mental health" was actually used. It appears that "rational-altruistic character" (Peck & Havighurst, 1960) is equivalent to "mentally healthy" for these judges. Of paramount significance are such traits as objectivity, rationality, insightfulness, and basic trust in other people. Self-confidence and appropriateness of behavior also are important components.

Table 4, on the other hand, reveals certain discrepancies among the raters' viewpoints. Raters A and D place less importance on richness of fantasy and on persistence than do Raters B and C. Rater B de-emphasizes interest in others, but stresses identification with children, as compared to the other raters. Rater C places less importance on expression of affection and on appropriateness of goals. Rater D does not place importance on decisiveness of judgment, or dominance, to the extent that the others do. It is nonetheless interesting to note that only one minor reversal of direction occurred in the correlation coefficients.

DISCUSSION

Even where judges largely agree, as these did (the reliability of their averaged ratings was .85), the construct system of each judge still shows certain unique aspects. The method of analyzing ratings which was used here reveals each judge's system of beliefs about the structure of human nature.

If, in some other study, judges were to disagree notably, such a method of analysis would identify even more sharply each judge's unique set of concepts. As a study in the psychological beliefs of people like professors, this might be interesting in itself. Beyond

this, however, clear, precise identification of the specific points at which differences exist could conceivably lead to clearer, more conscious, and more rational discussion among judges, until consensus was reached or they agree to disagree. In either case, this kind of analysis should prove a healthy reminder that the arguments are not necessarily about the facts of human nature, but may sometimes arise from differing individual construct systems. The construct systems, in this case, are laid out on the table, for all to see.

Such a procedure, needless to say, is not limited to studies of personality. It could as usefully be applied to studies of ethical values, beliefs about educational practices—indeed, any issue important enough to warrant some planned scaling of judgments, and a moderate amount of computer time.

Summary

A generalized method was developed for identifying the individual construct systems of judges of human behavior. The method is that of separately factor analyzing each judge's ratings on whatever scales are used in an assessment of people. Thereafter, correlation of the factors generated by several judges indicates the nature and degree of similarities and differences in the way the different judges define "human nature," and the way in which they believe different aspects of behavior are related to one another.

In the present instance, four educational psychologists rated 53 sophomore coeds, on 50 personality variables and on a scale of "overall mental health." The ratings were based on qualitative analysis of a battery including a biographical form, the Peck Sentence Completion (2-D), group TAT, Bown School Situations Analyzer, and an Educational Issues Questionnaire. The average interjudge correlation for all ratings was .85. Even so, some significant differences were found in the construct systems of the four judges. Each showed certain unique concepts about the structure of personality. Each differed at certain points, about the personality correlates of mental health, although there was agreement among all the judges that 12 personality variables are closely related to mental health.

The general logic of this procedure can be applied to *any* kind of judgments, regardless of the specific variables involved. It thus offers a generally applicable way of measuring individual construct systems, it appears, in *any* field where human beings make judgments. To differentiate it from *R*, *P*, and *Q* techniques, perhaps it might be called "*J* technique."

REFERENCES

Kelly, G. A. *The psychology of personal constructs.* New York: Norton, 1955. 2 vols.

Peck, R. F. Personality patterns of prospective teachers. *J. exp. Educ.* 1960, **29**, 169–175.

Peck, R. F., & Havighurst, R. J. *The psychology of character development.* New York: Wiley, 1960.

(Received June 24, 1961)

GRADUATE EDUCATION IN PSYCHOLOGY

Report of the Conference on Graduate Education in Psychology, sponsored by the Education and Training Board of the American Psychological Association and supported by a grant from the National Institute of Mental Health, U. S. Public Health Service; held at Miami Beach, Florida, November 29 to December 7, 1958.

Prepared by the Editorial Committee:
Anne Roe, *Chairman*,
and
John W. Gustad, Bruce V. Moore,
Sherman Ross, and Marie Skodak

Price $1.50

THE AMERICAN
PSYCHOLOGICAL ASSOCIATION
Dept. Grad
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Psychology and Mental Health

A report of the Institute on Education and Training for Psychological Contributions to Mental Health, held at Stanford University in August, 1955.

Edited by CHARLES R. STROTHER

Topics discussed at the Institute:

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: $1.75

Order from:

American Psychological Association, Inc.
1333 Sixteenth Street, N. W.
Washington 6, D. C.

New Reprint Now Available!!

The Journal of Educational Psychology

✓ REPRINTED VOLUMES, 1–10, 1910–1919
Cloth bound set..... $180.00
Paper bound set..... 165.00
Single vols., cloth bound..... 18.00
Single vols., paper bound..... 16.50

✓ VOLUMES 11–20, 1920–1929, are planned for reprint in the Spring of 1962, and advance orders are presently being accepted.

✓ COMPLETE SETS, Volumes 1–51, 1910–1960, with the aid of reprinted vols. 1–10, are available in limited quantities, bound or unbound. Please inquire.

✓ MANY ADDITIONAL VOLUMES, shorter runs, and single numbers are in stock. Please let us know your detailed wants.

Address All Orders And Inquiries to:

A. M. S. REPRINT COMPANY

56 EAST 13th STREET DEPT. E NEW YORK 3, N. Y.

(Subsidiary of Abrahams Magazine Service, Inc.)

ROGERS

Mental Hygiene in Elementary Education

DOROTHY ROGERS
State University of New York,
College of Education at Oswego

497 pages, 1957, $5.95

McKINNEY

Counseling for Personal Adjustment
In Schools and Colleges

FRED McKINNEY
University of Missouri

584 pages, 1958, $6.75

Houghton Mifflin
BOSTON · NEW YORK · ATLANTA
GENEVA, ILL. · DALLAS · PALO ALTO